



University of Groningen

## Full Implementation and Belief Restrictions

Ollar, Mariann; Penta, Antonio

*Published in:*  
American Economic Review

*DOI:*  
[10.1257/aer.20151462](https://doi.org/10.1257/aer.20151462)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Ollar, M., & Penta, A. (2017). Full Implementation and Belief Restrictions. *American Economic Review*, 107(8), 2243-2277. <https://doi.org/10.1257/aer.20151462>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Full Implementation and Belief Restrictions<sup>†</sup>

By MARIANN OLLÁR AND ANTONIO PENTA\*

*Multiplicity of equilibria and the dependence on strong common knowledge assumptions are well-known problems in mechanism design. We address them by studying full implementation via transfer schemes, under general restrictions on agents' beliefs. We show that incentive-compatible transfers ensure uniqueness—and hence full implementation—if they induce sufficiently weak strategic externalities. We then design transfers for full implementation by using information on beliefs in order to weaken the strategic externalities of the baseline canonical transfers. Our results rely on minimal restrictions on agents' beliefs, specifically on moments of the distribution of types, that arise naturally in applications. (JEL D62, D82, D83)*

The problem of multiplicity is a key concern for the design of real-world mechanisms and institutions. Unless all the solutions of a mechanism are consistent with the outcome which the designer wishes to implement, the designer may not confidently assume that the proposed mechanism will perform well. This is a well-known criticism of the widespread *partial implementation* approach to mechanism design, which requires only that there exists *one* strategy profile consistent with the chosen solution concept that guarantees desirable outcomes. The *full implementation* approach (cf. Maskin 1999) overcomes the problem of multiplicity, but in pursuit of greater generality, the existing literature has typically adopted rather complicated mechanisms.<sup>1</sup> Thus, while it addresses an important practical concern, the full implementation literature overall has provided limited insight into how real-world institutions could be designed to avoid the problem of multiplicity.

Another well-known limitation of the classical approach is its excessive reliance on common knowledge assumptions. This criticism, often referred to as the *Wilson*

\* Ollár: Department Econometrics and Finance, University of Groningen, Nettelbosje 2, 9747 AE Groningen (email: m.ollar@rug.nl); Penta: Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706 (email: apenta@ssc.wisc.edu). We are particularly grateful to the anonymous referees, whose comments greatly improved the paper. Special thanks also go to Larbi Alaoui, Ken Hendricks, Philippe Jehiel, George Mailath, Laurent Mathevet, Meg Meyer, Stephen Morris, Andy Postlewaite, Rakesh Vohra, and Bill Sandholm. We also thank seminar audiences at Stanford, NYU, UPenn, Oxford, Cambridge, Minnesota, UW-Madison, UPF, Ohio State, Georgetown, Bocconi, UCL, Queen Mary, Groningen, and at several conferences. Mariann Ollár is grateful for the financial support of the Warren Center for Network and Data Sciences at the University of Pennsylvania. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

<sup>†</sup> Go to <https://doi.org/10.1257/aer.20151462> to visit the article page for additional materials and author disclosure statement(s).

<sup>1</sup> See Jackson (1992) for an influential criticism of the tail-chasing mechanisms typically used in this literature.

*doctrine*, has recently received considerable attention in the literature on robust implementation. It is fair to say, however, that the aims of the Wilson doctrine, “[...] to conduct useful analyses of practical problems [...]” (Wilson 1987, p. 34), are still far from being fulfilled. In our view, this is due to two main reasons. First, most of this literature has focused on environments in which the designer has no information about the agents’ beliefs.<sup>2</sup> This extreme assumption represents a useful benchmark to address foundational questions, but significantly limits the possible applications of the theory to practical problems of mechanism design. Second, as far as full implementation is concerned, the literature has focused on characterization results which offer little insight on the properties that more realistic mechanisms should satisfy, in order to ensure full implementation. In this paper we address these points by pursuing a more pragmatic approach to full implementation, based on transfer schemes that only elicit agents’ payoff-relevant information, and relying on more realistic assumptions of common knowledge, intermediate between the classical and the belief-free approaches.

For the sake of illustration, consider the problem of efficient implementation. In environments with single-crossing preferences, the generalized VCG transfers of Crémer and McLean (1985) guarantee partial implementation of the efficient allocation in an ex post equilibrium, with essentially no restrictions on the strength of the preference interdependence. Hence, independent of the agents’ beliefs, truthful revelation (hence efficiency) is always achievable as part of a Bayes-Nash equilibrium (Bergemann and Morris 2005). The problem with this mechanism is that it typically admits also inefficient equilibria, which can be ruled out if and only if the interdependence in agents’ valuations is not too strong (Bergemann and Morris 2009a). But since in many cases preference interdependence is strong, this characterization is often regarded as a negative result.

In this paper we shift the focus of the analysis from preference interdependence to the *strategic externalities* in the mechanism, which—unlike preferences—can be affected by the designer. The problem with the VCG transfers, for instance, is that when agents’ preferences exhibit strong interdependence, the strategic externalities in the mechanism are strong, in that agents’ best responses are strongly affected by others’ strategies. This in turn generates multiplicity of equilibria, and hence failure of full implementation. But if the designer has some information about the agents’ beliefs, then preferences and strategic externalities need not be aligned; the strategic externalities can be weakened, so as to ensure uniqueness, even if preference interdependence is strong. To ensure that the unique solution implements the designer’s objective, the strategic externalities should be weakened in a way that preserves incentive compatibility—if not in the ex post sense, then at least for the beliefs consistent with the designer’s information. Note that this argument also suggests a tension between the robustness of the *partial* implementation result (achieved by the VCG mechanism in an ex post equilibrium) and the possibility of achieving *full* implementation (which, if preference interdependence is strong, necessarily requires information about beliefs).

<sup>2</sup>On the belief-free approach, see Bergemann and Morris (2005, 2009a, b, 2011) for static mechanisms and Müller (2016) and Penta (2015) for dynamic ones. We discuss the related literature in Section V.

Our model covers implementation problems with one-dimensional types, smooth allocation rules, and smooth valuation functions under varying assumptions on agents' beliefs. For this reason, we adopt a solution concept which extends rationalizability to environments with incomplete information and general assumptions on beliefs.<sup>3</sup> The resulting notion of implementation provides a unified framework to study full implementation under a broad class of *belief restrictions*, thereby allowing for varying degrees of robustness. It also formalizes the idea that the robustness of a mechanism is determined contextually with its design, and as such it can be chosen by the designer the same way that transfers are. This change in perspective allows us to move beyond the existing characterization results, to gain insights on what can still be achieved when the conditions for belief-free implementation are not met.

The general analysis parallels the example above. First we derive the canonical transfers, a generalization of well-known necessary conditions for ex post incentive-compatible payment schemes. Depending on the environment, and particularly on the strength of the preference interdependence, the canonical transfers may induce overly strong strategic externalities, which are problematic for full implementation. The second part of our design then exploits the belief restrictions to reduce the strategic externalities, so as to induce uniqueness. The conditions that guarantee full implementation relate the strength of the preference interdependence to the designer's information on agents' beliefs. This information takes the form of *moment conditions*, which represent weak restrictions on agents' beliefs and which arise naturally in applications.

Our results suggest a simple design strategy: start with the canonical transfers, and then compensate each agent for the strategic externality he faces, given everybody's reports. To deter agents from misreporting their types in order to inflate their compensation, each agent  $i$  is also asked to pay a fee equal to the expected cumulative marginal compensation, given his report:

$$t_i(m) = \underbrace{t_i^*(m)}_{\substack{\text{canonical transfers} \\ \text{(e.g., VCG)}}} + \underbrace{CSE_i(m_i, m_{-i})}_{\substack{\text{compensation for} \\ \text{strategic externality} \\ \text{(depends on everybody's report)}}} - \underbrace{\int^{m_i} E\left(\frac{\partial CSE_i}{\partial m_i} \middle| s_i\right) ds_i}_{\substack{\text{belief-based adjustment: cumulative} \\ \text{expected marginal compensation} \\ \text{(only depends on } i\text{'s report)}}}.$$

The first term we add to the canonical transfers reduces the strategic externalities and ensures uniqueness; the last term, derived from the designer's information about beliefs, restores incentive compatibility. Full implementation follows.

As applications of our main results, we study smooth environments that satisfy standard single-crossing properties and a *public concavity* condition, which generalizes important classes of models in the literature. Under these restrictions on preferences, we show that: (i) in the Bayesian environments that are common in the classical and applied literature, full implementation via transfers is always possible if types are independent, or if they are affiliated and valuations are supermodular,

<sup>3</sup>Formally, our solution concept is a special case of Battigalli and Siniscalchi's (2003)  $\Delta$ -Rationalizability and encompasses several notions such as belief-free (Bergemann and Morris 2009a) and interim correlated rationalizability (Dekel, Fudenberg, and Morris 2007). Further connections are discussed in Sections II and V.

regardless of preference interdependence; (ii) within these settings, ex post incentive compatibility is possible if and only if (interim) dominant-strategy implementation is; (iii) in non-Bayesian environments, in which only the conditional averages of types are common knowledge, implementation can always be achieved, provided that the conditional averages of the opponents' types are constant or increasing in an agent's own type.

Finally, we show that mechanisms with weak strategic externalities have further desirable properties, such as low sensitivity to misspecifications of agents' beliefs. This result suggests further notions of robustness as well as a novel concept of approximate implementation.

The rest of the paper is organized as follows. Section I introduces the model and the leading examples. Section II presents the notion of implementation. Section III provides the main results on full implementation via transfers. Section IV contains the applications and the sensitivity analysis. The related literature is discussed in Section V. Section VI concludes.

## I. Model

*Environments and Mechanisms.*—We consider environments with transferable utility with a finite set of agents  $I = \{1, \dots, n\}$ , in which the space of allocations  $X$  is a compact and convex subset of a Euclidean space. Agents privately observe their payoff types  $\theta_i \in \Theta_i := [\underline{\theta}_i, \bar{\theta}_i] \subseteq \mathbb{R}$ , and we adopt the standard notation  $\theta_{-i} \in \Theta_{-i} = \times_{j \neq i} \Theta_j$  and  $\theta \in \Theta = \times_{i \in I} \Theta_i$  for profiles. Agent  $i$ 's *valuation function* is  $v_i : X \times \Theta \rightarrow \mathbb{R}$ , assumed three times continuously differentiable, and we let  $t_i \in \mathbb{R}$  denote the private transfer to agent  $i$ : for each outcome  $(x, \theta, (t_i)_{i \in I})$ ,  $i$ 's utility is equal to  $v_i(x, \theta) + t_i$ . The tuple  $\langle I, (\Theta_i, v_i)_{i \in I} \rangle$  is common knowledge among the agents. If  $v_i$  is constant in  $\theta_{-i}$  for every  $i$ , then the environment has *private values*. If not, it has *interdependent values*.

An *allocation rule* is a mapping  $d : \Theta \rightarrow X$  which assigns to each payoff state the allocation that the designer wishes to implement. We focus on allocation rules that are twice continuously differentiable and *responsive*, in the sense that for all  $i$  and  $\theta_i \neq \theta'_i$ , there exists  $\theta_{-i} \in \Theta_{-i}$  such that  $d(\theta_i, \theta_{-i}) \neq d(\theta'_i, \theta_{-i})$  (e.g., Bergemann and Morris 2009a).

The model accommodates general externalities in consumption, including both pure cases of private and public divisible goods. The main substantive restrictions are the one-dimensionality of types and the smoothness of the allocation function, which for instance rules out standard auction applications. We will use the notation  $\partial f / \partial x$  for all derivatives, with the understanding that when  $X$  is multidimensional,  $\frac{\partial v_i}{\partial x}(x, \theta)$  and  $\frac{\partial d}{\partial \theta_i}(\theta)$  denote the vectors of partial derivatives and  $\frac{\partial v_i}{\partial x}(x, \theta) \cdot \frac{\partial d}{\partial \theta_i}(\theta)$  denotes their inner product.

We consider direct mechanisms, in which agents report their type and the allocation is chosen according to  $d$ . A direct mechanism is thus uniquely determined by a *transfer scheme*  $t = (t_i)_{i \in I}$ ,  $t_i : \Theta \rightarrow \mathbb{R}$ , which specifies the transfer to each agent  $i$ , for all profiles of reports  $m \in \Theta$ .<sup>4</sup> (To distinguish the report

<sup>4</sup>The full implementation literature typically focuses on characterizations of the implementable  $f : \Theta \rightarrow Y$ , where  $Y$  denotes the space of outcomes (see, e.g., Bergemann and Morris 2009a). For  $Y = X \times \mathbb{R}^n$ , such

from the state, we maintain the notation  $m_i$  even though the message spaces are  $M_i = \Theta_i$ .) We focus on transfer schemes that are twice continuously differentiable and bounded. Thus, under the maintained assumptions, a transfer scheme induces a game with ex post payoff functions  $U_i(m; \theta) = v_i(d(m), \theta) + t_i(m)$  that are twice continuously differentiable and bounded.<sup>5</sup> For every  $\theta_i \in \Theta_i$ ,  $\mu \in \Delta(M_{-i} \times \Theta_{-i})$ , and  $m_i \in M_i$ , we let  $EU_{\theta_i}^\mu(m_i) = \int_{M_{-i} \times \Theta_{-i}} U_i(m_i, m_{-i}; \theta_i, \theta_{-i}) d\mu$  denote agent  $i$ 's expected payoff from message  $m_i$ , if  $i$ 's type is  $\theta_i$  and his conjectures are  $\mu$ , and define  $BR_{\theta_i}(\mu) := \arg \max_{m_i \in M_i} EU_{\theta_i}^\mu(m_i)$ .

*Belief Restrictions.*—We model belief restrictions as sets of possible beliefs for each type of every agent. Formally, the *belief restrictions* are a commonly known collection  $\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$  such that  $B_{\theta_i} \subseteq \Delta(\Theta_{-i})$  is nonempty and convex for all  $i$  and  $\theta_i$ , and  $B_i : \theta_i \mapsto B_{\theta_i} \subseteq \Delta(\Theta_{-i})$  is continuous for every  $i$ . If  $\mathcal{B}$  and  $\mathcal{B}'$  are such that  $B_{\theta_i} \subseteq B'_{\theta_i}$  for all  $\theta_i$  and  $i$ , we write  $\mathcal{B} \subseteq \mathcal{B}'$ .

This formulation is fairly general. For instance, if  $B_{\theta_i}$  is a singleton for every  $\theta_i$  and  $i$ , then we obtain a standard Bayesian environment, in which agents' hierarchies of beliefs are uniquely pinned down by their payoff types. The further special case of a common prior model requires that  $B_{\theta_i} = \{b_{\theta_i}\}$  are such that there exists  $p \in \Delta(\Theta)$  such that  $b_{\theta_i} = p(\cdot | \theta_i) \in \Delta(\Theta_{-i})$  for each  $i$  and  $\theta_i$ . If, furthermore,  $B_{\theta_i} = B_{\theta'_i}$  for all  $i$  and all  $\theta_i, \theta'_i \in \Theta_i$ , then we obtain the case of independent types (cf. Example 1). At the opposite extreme, if  $B_{\theta_i} = \Delta(\Theta_{-i})$  for all  $\theta_i$  and  $i$ , then there are essentially no restrictions on beliefs (beyond their support, that is), and the model coincides with the belief-free environments that are common in the literature on robust mechanism design (see footnote 2). Such vacuous restrictions are thus denoted by  $\mathcal{B}^{BF}$ . Our model also accommodates settings, intermediate between the Bayesian and belief-free cases, in which some restrictions on beliefs are maintained but not to the point that belief hierarchies are uniquely determined by the payoff types. In those cases,  $\mathcal{B}$  represents the designer's partial information about agents' beliefs. If  $\mathcal{B} \subseteq \mathcal{B}'$ , then  $\mathcal{B}'$  entails weaker restrictions than  $\mathcal{B}$ .

### A. Leading Examples

**Example 1 (Full Implementation in a Common Prior Model):** Consider an environment with two agents,  $i \in \{1, 2\}$ . The social planner chooses a quantity  $x \in X \subseteq \mathbb{R}_+$  of a public good, with cost of production  $c(x) = \frac{1}{2}x^2$ . Agents' valuation functions are  $v_i(x, \theta) = (\theta_i + \gamma\theta_j)x$ , where  $\gamma \geq 0$  is a parameter of preference interdependence: if  $\gamma = 0$ , this is a private-value setting; if  $\gamma > 0$ , values are interdependent. The planner knows that types are i.i.d. draws from a uniform

---

characterization results can be used to check whether a given  $f(\cdot) = (d(\cdot), t(\cdot))$  is implementable by a direct mechanism (and hence whether a given transfer scheme implements  $d$ ), but do not provide insights on *how to design* transfers for full implementation. Since we are interested in this kind of constructive insights, we maintain here the standard setup of the partial implementation literature: that is, we only take  $d : \Theta \rightarrow X$  as given, and let the designer choose  $t : \Theta \rightarrow \mathbb{R}^n$ . The restriction to *direct mechanisms* also entails some loss of generality for full implementation, but in these environments it allows an easier comparison with the partial implementation literature, by making transparent what features of an incentive-compatible transfer scheme may or may not be problematic for full implementation.

<sup>5</sup>Since  $(d, t)$  will be clear from the context, we don't emphasize the dependence of the payoff functions on  $(d, t)$ . Also, for any measurable set  $E$ , we let  $\Delta(E)$  denote the set of probability measures on its Borel sigma-algebra.



distribution over  $\Theta_i \equiv [0, 1]$ , denoted by  $\bar{v}_{\Theta_i}$ , and that this is common knowledge among the agents. This is a standard common prior environment, with independently distributed types and interdependent values. The planner's information about agents' beliefs is represented by belief restrictions  $\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$  such that  $B_{\theta_i} = \{\bar{v}_{\Theta_j}\}$  for every  $i, j \neq i$ , and  $\theta_i \in \Theta_i$ .

The social planner wishes to implement the efficient level of public good,  $d(\theta) = (1 + \gamma)(\theta_1 + \theta_2)$ . This allocation rule can be partially implemented by the generalized VCG transfers

$$(1) \quad t_i^{VCG}(m) = -(1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i m_j \right).$$

Given this, for any pair  $(m_j, \theta_j)$  of  $j$ 's report and type, the ex post best-reply for type  $\theta_i$  is

$$(2) \quad BR_{\theta_i}^{VCG}(m_j, \theta_j) = \text{proj}_{[0, 1]}(\theta_i + \gamma(\theta_j - m_j)).^6$$

Observe that, for any  $\gamma \geq 0$ , truthful revelation ( $m_i(\theta_i) = \theta_i$ ) is a best response to the opponent's truthful strategy ( $m_j(\theta_j) = \theta_j$ ), and hence the efficient allocation rule is *partially implemented* independent of agents' beliefs. Furthermore, if  $\gamma < 1$ , equation (2) is a contraction, and its iteration delivers truthful revelation as the only rationalizable strategy. In this case, the VCG mechanism also guarantees belief-free *full* implementation (Bergemann and Morris 2009a). But full implementation fails if  $\gamma \geq 1$ . (In the symmetric case with  $n$  agents, it can be shown that no mechanism achieves belief-free full implementation if  $\gamma \geq 1/(n - 1)$ .)

Hence, with weak interdependence in valuations, the designer need not rely on the common prior: the VCG mechanism ensures full implementation in the belief-free model  $\mathcal{B}^{BF} \supset \mathcal{B}$ . If the interdependence is strong, however, full implementation fails, even under the  $\mathcal{B}$ -restrictions. For instance, if  $\gamma = 2$  and types are independently and uniformly distributed, the strategy profile  $(\hat{m}_1(\theta_1) = 1, \hat{m}_2(\theta_2) = 0)$  is also a Bayes-Nash equilibrium of this mechanism, and it is inefficient.

Being designed to achieve ex post incentive compatibility, the VCG mechanism ignores any information about agents' beliefs. We propose next a different set of transfers, which do exploit some information contained in the common prior (namely, that  $E(\theta_j | \theta_i) = 0.5$  for all  $\theta_i$  and  $i$ ):

$$(3) \quad t_i^\rho(m) := -(1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i E(\theta_j | \theta_i) \right) = -(1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i \cdot 0.5 \right).$$

These transfers induce the following best response function:

$$(4) \quad BR_{\theta_i}^\rho(\hat{m}_j(\cdot)) = \text{proj}_{[0, 1]}(\theta_i + \gamma[E(\theta_j | \theta_i) - 0.5]).$$

Since, under the common prior,  $E(\theta_j | \theta_i) = 0.5$  for all  $\theta_i$ , the term in square brackets cancels out for all types. Truthful revelation therefore is *strictly dominant*, regardless

<sup>6</sup>For any  $y \in \mathbb{R}$ ,  $\text{proj}_{[0, 1]}(y) := \arg \min_{m_i \in [0, 1]} |m_i - y|$  denotes the projection of  $y$  on the interval  $[0, 1]$ .

of the strength of preference interdependence,  $\gamma$ . Note that this holds for all beliefs that satisfy the *moment condition*  $E(\theta_j | \theta_i) = 0.5$  for all  $\theta_i$ . Hence, full implementation is guaranteed not just for the common prior model,  $\mathcal{B}$ , but also for the weaker restrictions  $\hat{\mathcal{B}} = ((\hat{B}_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$  defined as  $\hat{B}_{\theta_i} := \{b_i \in \Delta(\Theta_j) : \int \theta_j \cdot db_i = 0.5\}$ . Moreover, since truthful revelation is dominant in this mechanism, given  $\hat{\mathcal{B}}$ , such restrictions need not be common knowledge among the agents: as long as  $E(\theta_j | \theta_i) = 0.5$  for all  $\theta_i$ , full implementation obtains independent of higher-order beliefs.

The previous example poses a standard Bayesian implementation problem, in which the planner's information is represented by a common prior model,  $\mathcal{B}$ . Full implementation, however, need not rely on the full strength of these assumptions. If  $0 \leq \gamma < 1$ , the VCG mechanism ensures belief-free implementation, that is for all beliefs consistent with  $\mathcal{B}^{BF} \supset \mathcal{B}$ . If  $\gamma \geq 1$ , the transfers in (3) achieve full implementation for all beliefs consistent with  $\hat{\mathcal{B}} \supset \mathcal{B}$ . Clearly, the precise definition of  $\hat{\mathcal{B}}$  depends on the particular *moment condition* we used to design the transfers. Had we used a different condition, full implementation might have obtained for different belief restrictions  $\mathcal{B}' \supset \mathcal{B}$ . Thus, it is not only true that  $\mathcal{B}$ , which represents the designer's information, need not coincide with the set of beliefs for which implementation is ensured (such as  $\mathcal{B}^{BF}$  or  $\hat{\mathcal{B}}$  in the example), but the latter set is itself determined by the planner's choice of the mechanism.

In Section II we introduce the notion of implementation, and formalize the sense in which the strength of the strategic externalities, not of the preference interdependence, is key for full implementation. The two go hand in hand in belief-free environments, but need not coincide if the designer has some information about agents' beliefs. In Section III we develop a design principle which consists of using properly chosen belief restrictions to weaken the strategic externalities of a baseline canonical mechanism. We show that *moment conditions*, formally introduced in Section IB, are particularly suited to this task.

In the example above, a moment condition enabled us to completely offset the strategic externalities of the VCG mechanism, thereby ensuring full implementation in dominant strategies. In the general case in which strategic externalities cannot be completely eliminated, our design strategy pursues contractive best replies, to ensure that truthful revelation is the unique rationalizable outcome. The next example illustrates this point in a non-Bayesian model.

**Example 2 (Full Implementation without a Common Prior):** Consider an environment with three agents,  $i \in \{1, 2, 3\}$ , who commonly believe that types  $\theta_i \in [0, 1]$  are i.i.d. draws from some distribution  $\Phi$ . The distribution itself, however, is not necessarily known by the agents, and most importantly it is *unknown* to the designer. This environment therefore provides an example both of non-Bayesian belief restrictions and of a situation in which the designer may know *less* than what is commonly known by the agents.

Preferences are such that  $v_i(x, \theta) = (\theta_i + \gamma\theta_j + \delta\theta_k)x$  for each  $i$ , where  $x \in \mathbb{R}_+$  denotes the quantity of public good,  $\gamma, \delta \in \mathbb{R}$ , and where we let  $j := i + 1 \pmod{3}$  and  $k := i + 2 \pmod{3}$ . If the cost of production is the same as in the previous example, the efficient allocation rule is  $d(\theta) = \kappa(\theta_1 + \theta_2 + \theta_3)$  where



$\kappa \equiv (1 + \gamma + \delta)$ . The VCG transfers are  $t_i^{VCG}(m) = -\kappa(0.5m_i^2 + m_i(\gamma m_j + \delta m_k))$ , which induce the following interim best reply:

$$BR_{\theta_i}^{VCG} = \text{proj}_{[0,1]}(\theta_i + E(\gamma(\theta_j - m_j) + \delta(\theta_k - m_k) | \theta_i)).$$

Now, suppose that  $\gamma = 4/3$  and  $\delta = -2/3$ . With these parameter values, any report profile is rationalizable, and belief-free implementation fails. The following transfers instead achieve full implementation:  $t_i^\rho(m) = t_i^{VCG}(m) + m_i \kappa \gamma (m_j - m_k)$ . With these transfers, the best reply is

$$\begin{aligned} (5) \quad BR_{\theta_i}^\rho &= \text{proj}_{[0,1]}(\theta_i + \gamma E(\theta_j - \theta_k | \theta_i) + (\gamma + \delta) E(\theta_k - m_k | \theta_i)) \\ &= \text{proj}_{[0,1]}(\theta_i + (\gamma + \delta) E(\theta_k - m_k | \theta_i)). \end{aligned}$$

The simplification in the second line occurs because, regardless of the distribution  $\Phi$ , we have that  $E(\theta_j - \theta_k | \theta_i) = 0$  in this model. Unlike the previous example, strategic externalities are not eliminated in this case. However, for the values of parameters specified above, the term  $(\gamma + \delta) = 2/3 < 1$ . Hence, the best replies induce a contraction, which delivers truthful revelation as the only rationalizable profile. Similar to the previous example, full implementation only relies on common knowledge of the *moment condition*  $E(\theta_j - \theta_k | \theta_i) = 0$  for all  $\theta_i$ . Formally, the belief restrictions  $\mathcal{B}$  in this model are such that  $B_{\theta_i} = \{b_i \in \Delta(\Theta_{-i}) : \exists \phi \in \Delta([0,1]) \text{ such that } b_i = \otimes_{j \neq i} \phi\}$ , whereas transfers  $t^\rho$  achieve full implementation for the weaker restrictions  $\mathcal{B}' \supseteq \mathcal{B}$ , such that  $B'_{\theta_i} = \{b_i \in \Delta(\Theta_{-i}) : \int (\theta_k - \theta_j) db_i = 0\}$ , whenever  $|\gamma + \delta| < 1$ .

### B. Moment Conditions

As shown above, our design strategy exploits a special class of belief restrictions: moment conditions. In this section we introduce the concept formally.

**DEFINITION 1:** A  $\mathcal{B}$ -consistent *moment condition* is a collection  $\rho = (L_i, f_i)_{i \in I}$  of twice continuously differentiable functions,  $L_i : \Theta_{-i} \rightarrow \mathbb{R}$  and  $f_i : \Theta_i \rightarrow \mathbb{R}$ , such that given  $\mathcal{B}$  it is common knowledge that  $i$ 's expectation of  $L_i(\theta_{-i})$  varies with  $\theta_i$  according to  $f_i$ . We let  $\varrho(\mathcal{B})$  denote the set of moment conditions that are consistent with  $\mathcal{B}$ . Formally,  $(L_i, f_i)_{i \in I} \in \varrho(\mathcal{B})$  if and only if

$$(6) \quad \int_{\Theta_{-i}} L_i(\theta_{-i}) db_i = f_i(\theta_i) \quad \text{for all } i, \theta_i, \text{ and } b_i \in B_{\theta_i}.$$

For each  $\mathcal{B}$ -consistent moment condition,  $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B})$ , we define the belief restrictions in which only common knowledge of  $\rho$  is maintained, as  $\mathcal{B}^\rho = ((B_{\theta_i}^\rho)_{\theta_i \in \Theta_i})_{i \in I}$  such that

$$B_{\theta_i}^\rho := \left\{ b_i \in \Delta(\Theta_{-i}) : \int_{\Theta_{-i}} L_i(\theta_{-i}) db_i = f_i(\theta_i) \right\} \quad \text{for all } i \text{ and } \theta_i.$$

It is easy to see that, for any  $\mathcal{B}$  and  $\rho \in \varrho(\mathcal{B})$ ,  $\mathcal{B}^\rho$  entails weaker restrictions than  $\mathcal{B}$  (that is,  $\mathcal{B}^\rho \supseteq \mathcal{B}$ ). The next two examples show how moment conditions are implicit in standard models.

**Example 3 (Unobserved Heterogeneity and Fundamental Value Models):** Suppose that types  $\theta_i$  are i.i.d. draws from a distribution  $F_\eta$ , where  $\eta$  is drawn from another distribution  $H$  and is unobserved by the designer. This model entails many moment conditions. For instance, it is common knowledge in this model that  $E(\theta_l - \theta_k | \theta_i) = 0$  for all  $\theta_i$  and  $i \neq l, k$ . This is represented by setting  $L_i(\theta_{-i}) = \theta_l - \theta_k$  for some  $l, k \neq i$  and  $f_i(\theta_i) = 0$  for all  $\theta_i$ . (This moment condition was used in Example 2.) Notice that this is the case regardless of the details of the distributions  $H$  and  $F_\eta$ , and on whether  $\eta$  is observed by the agents. Examples of the first case include *models of unobserved heterogeneity* (e.g., Aradillas-López, Gandhi, and Quint 2013). Examples of the second case are provided by *fundamental value models*, in which  $\theta_i = \theta_0 + \varepsilon_i$ , where  $\varepsilon_i$ s are i.i.d. across agents and independent of  $\theta_0$ , which in turn is drawn from a normal distribution but remains unobserved (e.g., Grossman and Stiglitz 1980 and Hellwig 1980).

**Example 4 (Spatial Values):** Consider an environment with two groups of agents (e.g., distinct by geographic location, technology, etc.). Agents are assigned to group 1 independently with probability  $p$ , and they inherit the type of their group, drawn independently from a distribution with mean  $E(\theta)$ . An agent's type is his private information, his group is known to the designer but not to the other agents (e.g., Ausubel and Baranov 2013). In this model it is common knowledge that  $E(\theta_j | \theta_i) = p(i) \theta_i + (1 - p(i)) E(\theta)$ , where  $p(i) = p$  if  $i$  belongs to group 1, and  $(1 - p)$  otherwise. The corresponding moment condition obtains setting  $L_i(\theta_{-i}) = \theta_j$  for some  $j \neq i$  and  $f_i(\theta_i) = p(i) \theta_i + (1 - p(i)) E(\theta)$ .

Moment conditions arise naturally in many settings, in which knowledge of some moments of the distribution is a more basic and realistic kind of information than the one implicit in standard common prior models. Consider the following examples.

**Example 5 (Uncorrelated Types without a Prior):** Suppose that the designer has data showing no significant correlations across agents. His information, however, does not include the entire distribution of agents' types, but only some moments  $\rho$  of that distribution. In this case, the designer's information itself consists of moment conditions (that is,  $\mathcal{B} = \mathcal{B}^\rho$ ). For example, if types are uncorrelated, for each  $i, j$ , and  $\theta_i$ , we have  $E(\theta_j | \theta_i) = y_j$  for some  $y_j \in \mathbb{R}$ . In this case, a moment condition obtains by letting  $L_i(\theta_{-i}) = \theta_j$  and  $f_i(\theta_i) = y_j$ .

**Example 6 (Estimation-Based Conditions):** Consider a situation in which past data facilitate conditional predictions of agents' types in the form of linear regressions. Linear regressions are moment conditions, with  $L_i(\theta_{-i}) = \theta_j$  for  $j \neq i$  and  $f_i(\theta_i) = \hat{c}_i + \hat{a}_i \theta_i$  (where  $\hat{a}_i$  and  $\hat{c}_i$  are the estimated coefficients). Alternatively, past data may only report aggregate statistics of the distributions, so that only conditional expectations of the average of types may be available. In this case, a moment condition is obtained by letting  $L_i(\theta_{-i}) = \frac{1}{n-1} \sum_{j \neq i} \theta_j$ , and so on.

Econometric methods often provide a description of the environment in terms of conditional moments of the distributions, rather than a single *common prior*. In these cases, the very belief-restrictions  $\mathcal{B}$  can be specified as the set of all beliefs consistent with such moment conditions, taken as a primitive. Examples 5 and 6 are instances of such situations.

Observe that, in general, *any* belief restriction entails common knowledge of some moment conditions (that is,  $\varrho(\mathcal{B}) \neq \emptyset$  for any  $\mathcal{B}$ ). At a minimum, condition (6) is satisfied for any constant functions  $\bar{L}_i(\cdot) = \bar{f}_i(\cdot) = \bar{y}$ . In a belief-free environment, only such trivial moment conditions are commonly known. (Conversely,  $\mathcal{B}^\rho \equiv \mathcal{B}^{BF}$  whenever  $\rho = (\bar{L}_i, \bar{f}_i)_{i \in I}$  consists of such trivial moment conditions.) In general, the stronger the belief-restrictions (i.e., the smaller the sets  $\mathcal{B}$ ), the richer the set of moment conditions:  $\varrho(\mathcal{B}') \subseteq \varrho(\mathcal{B})$  if  $\mathcal{B} \subseteq \mathcal{B}'$ . Hence, common prior models are maximal in the set of moment conditions they satisfy: if  $\mathcal{B}$  is a common prior model, *any* collection of functions  $L_i : \Theta_{-i} \rightarrow \mathbb{R}$  satisfies  $(L_i, f_i^L)_{i \in I} \in \varrho(\mathcal{B})$  for  $f_i^L(\theta_i) := \int_{\Theta_{-i}} L_i(\theta_{-i}) dp(\cdot | \theta_i)$ , and hence the designer has maximum freedom to choose a suitable moment condition (cf. Section IV).

## II. Implementation

Our solution concept,  $\mathcal{B}$ -rationalizability, is defined by an iterated deletion procedure in which, for each type  $\theta_i$ , a report survives the  $k$ th round of deletion if and only if it can be justified by conjectures (i.e., joint beliefs over the opponents' types and their behavior in the mechanism) that are consistent with the belief restrictions for that type, and with the previous rounds of deletion. For every  $i$  and  $\theta_i$ , the set of conjectures that are consistent with the belief restrictions for type  $\theta_i$  is defined as  $C_{\theta_i}^{\mathcal{B}} := \{\mu_i \in \Delta(M_{-i} \times \Theta_{-i}) : \text{marg}_{\Theta_{-i}} \mu_i \in B_{\theta_i}\}$ .

**DEFINITION 2 ( $\mathcal{B}$ -Rationalizability):** Fix a direct mechanism and belief restrictions  $\mathcal{B}$ . For every  $i \in I$ , let  $R_i^{\mathcal{B},0} = \Theta_i \times M_i$  and for each  $k = 1, 2, \dots$ , let

$$R_{-i}^{\mathcal{B},k-1} = \times_{j \neq i} R_j^{\mathcal{B},k-1},$$

$$R_i^{\mathcal{B},k} = \left\{ (\theta_i, m_i) : m_i \in BR_{\theta_i}(\mu_i) \text{ for some } \mu_i \in C_{\theta_i}^{\mathcal{B}} \cap \Delta(R_{-i}^{\mathcal{B},k-1}) \right\},$$

and  $R_i^{\mathcal{B}} = \bigcap_{k \geq 0} R_i^{\mathcal{B},k}$ .<sup>7</sup> The set of  $\mathcal{B}$ -rationalizable messages for type  $\theta_i$  is

$$R_i^{\mathcal{B}}(\theta_i) := \{m_i : (\theta_i, m_i) \in R_i^{\mathcal{B}}\}.$$

**DEFINITION 3 (Full Implementation):** Allocation rule  $d$  is  $\mathcal{B}$ -implemented by transfer scheme  $t = (t_i)_{i \in I}$ , if truthful revelation is the only  $\mathcal{B}$ -rationalizable strategy profile in the direct mechanism  $(d, t)$  (that is, if  $R_i^{\mathcal{B}}(\theta_i) = \{\theta_i\}$  for all  $i$  and  $\theta_i$ ).<sup>8</sup> If

<sup>7</sup>  $\mathcal{B}$ -rationalizability is a special case of Battigalli and Siniscalchi's (2003)  $\Delta$ -rationalizability, obtained by setting their  $\Delta$ -restrictions such that  $\Delta_{\theta_i} = C_{\theta_i}^{\mathcal{B}}$  for all  $\theta_i$ . Under the maintained assumptions of Section II, the sets  $R_i^{\mathcal{B},k-1}$  are measurable and well defined for the mechanisms we consider, for every  $k$ .

<sup>8</sup> A weaker notion of implementation would allow nontruthful reports, provided that they all induce the same allocation as the true type profile (i.e.,  $R^{\mathcal{B}}(\theta) \neq \emptyset$  and  $d(m) = d(\theta)$  for all  $m \in R^{\mathcal{B}}(\theta)$ ). But it can be shown that the two notions coincide for responsive allocation rules.

this occurs in only one round of deletion (i.e., if  $R_i^{\mathcal{B},1}(\theta_i) = \{\theta_i\}$  for all  $i$  and  $\theta_i$ ), then we say that  $t$  implements  $d$  in  $\mathcal{B}$ -dominant strategies.

We say that  $d$  is  $\mathcal{B}$ -implementable (respectively,  $\mathcal{B}$ -DS implementable) if there exist transfers that  $\mathcal{B}$ -implement  $d$  (respectively, implement  $d$  in  $\mathcal{B}$ -dominant strategies).

As the belief restrictions are varied,  $\mathcal{B}$ -rationalizability coincides with various versions of rationalizability, some of which play an important role in the literature on robustness and implementation (see Section V). Also note that  $\mathcal{B}$ -rationalizability is a weak solution concept, and full implementation results are stronger if obtained with respect to a weaker solution concept. Hence, sufficient conditions for full  $\mathcal{B}$ -implementation guarantee full implementation with respect to any (nonempty) refinement of  $\mathcal{B}$ -rationalizability. Finally, it can be shown that  $\mathcal{B}$ -rationalizability characterizes the set of all Bayes-Nash equilibrium strategies, taking the union over all type spaces that are consistent with  $\mathcal{B}$ . Full  $\mathcal{B}$ -implementation therefore can be seen as a shortcut to analyze standard questions of Bayesian implementation for general belief restrictions.

Note that  $\mathcal{B}$ -DS implementation is more demanding and *more robust* than  $\mathcal{B}$ -implementation. As shown in Example 1, if truthful implementation is achieved in one round of  $\mathcal{B}$ -rationalizability, then truthful revelation is the only best response to all conjectures consistent with  $\mathcal{B}$ . In this case, full implementation obtains independent of higher-order beliefs, so the belief restrictions need not be common knowledge among the agents.

It is immediate from Definition 3 that, in order to achieve  $\mathcal{B}$ -implementation, the truthful profile must be a mutual best response for every type and for all conjectures consistent with the belief restrictions. This suggests the following notion of incentive compatibility.

**DEFINITION 4:** A direct mechanism is strictly  $\mathcal{B}$ -incentive compatible ( $\mathcal{B}$ -IC) if, for every agent and every type, truthful revelation is a strict best response to all conjectures that are consistent with the belief restrictions and concentrated on the opponents' truthful profile. Formally, if  $BR_{\theta_i}(\mu) = \{\theta_i\}$  for all  $i \in I$ ,  $\theta_i \in \Theta_i$ , and for all  $\mu \in C_{\theta_i}^{\mathcal{B}}$  such that  $\mu(\{(\theta_{-i}, m_{-i}) : m_{-i} = \theta_{-i}\}) = 1$ .

It is easy to verify that strict  $\mathcal{B}$ -IC coincides with strict ex post incentive compatibility (EPIC) if  $\mathcal{B} = \mathcal{B}^{BF}$ , and with strict interim (or Bayesian) incentive compatibility if  $\mathcal{B}$  is a standard type space. The following results are straightforward, from Definitions 3 and 4.

**Remark 1:** (i) Strict  $\mathcal{B}$ -IC is a necessary condition for  $\mathcal{B}$ -implementation. (ii) If a direct mechanism is strictly  $\mathcal{B}'$ -IC, then it is strictly  $\mathcal{B}$ -IC for all stronger restrictions  $\mathcal{B} \subseteq \mathcal{B}'$ . (iii) If  $d$  is  $\mathcal{B}'$ -implementable, then it is  $\mathcal{B}$ -implementable for all stronger restrictions  $\mathcal{B} \subseteq \mathcal{B}'$ .

The last point formalizes the idea, discussed in Section IA, that achieving implementation with respect to  $\mathcal{B}$  (the beliefs consistent with the designer's information) is the minimum objective. The notion of implementation, however, implicitly

accounts for the possibility of achieving full implementation for weaker belief restrictions  $\mathcal{B}' \supseteq \mathcal{B}$ , which would ensure a more robust result. In Example 1, for instance, depending on the parameter  $\gamma$ , full implementation could be obtained with respect to  $\mathcal{B}^{BF}$  or  $\hat{\mathcal{B}}$ , both of which are weaker than the designer's information in that example. Hence, if  $d$  is  $\mathcal{B}^\rho$ -implementable for some  $\rho \in \varrho(\mathcal{B})$ ,  $\mathcal{B}$ -implementation is achieved in a *more robust* sense (that is, relying on weaker common knowledge assumptions, namely  $\mathcal{B}^\rho \supseteq \mathcal{B}$ ).

As usual, incentive compatibility does not suffice for full implementation. We provide next some sufficient conditions, which will inform the design of transfers in Section III.<sup>9</sup>

**THEOREM 1:** *Let  $(d, t)$  be strictly  $\mathcal{B}$ -incentive compatible, with twice continuously differentiable transfers  $t_i : M \rightarrow \mathbb{R}$  and such that, for every  $i \in I$ ,  $\theta_i \in \Theta_i$ , and  $\mu \in C_{\theta_i}^{\mathcal{B}}$ ,  $EU_{\theta_i}^\mu : M_i \rightarrow \mathbb{R}$  is strictly concave. Then,  $(d, t)$  achieves full  $\mathcal{B}$ -implementation if the following holds:*

( $\mathcal{B}$ -Limited Strategic Externalities ( $\mathcal{B}$ -LSE)) *for all  $i$  and  $\theta_i$ , for all  $\mu \in C_{\theta_i}^{\mathcal{B}}$  and  $m_i, m'_i \in M_i$ ,*

$$(7) \quad \left| \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial^2 m_i}(m'_i, m_{-i}; \theta_i, \theta_{-i}) d\mu \right| > \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j}(m_i, m_{-i}; \theta_i, \theta_{-i}) \right| d\mu.$$

To understand this result, consider the first-order condition of type  $\theta_i$ 's optimization problem, given conjectures  $\mu \in C_{\theta_i}^{\mathcal{B}}$ :  $\int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(m_i, m_{-i}; \theta_i, \theta_{-i}) d\mu = 0$ . Under the concavity assumption in the theorem, this condition is both necessary and sufficient for an interior  $m_i^*$  to be a best response to  $\mu \in C_{\theta_i}^{\mathcal{B}}$ . Then, the second derivative  $\frac{\partial^2 U_i}{\partial m_i \partial m_j}(m_i, m_{-i}; \theta_i, \theta_{-i})$  measures the effect of  $j$ 's report on  $i$ 's best response, and hence  $j$ 's *strategic externality* on  $i$ . Condition (7) requires the *own effect* to be stronger than the opponents' effects, considered jointly. This condition therefore captures the idea that strategic externalities should not be too large.

Theorem 1 extends a result from Moulin (1984), which ensured uniqueness in *nice games* with complete information. The proof of Theorem 1, however, requires a different argument. This is partly due to the infinite-dimensional strategy spaces, but also to the robustness requirement implicit in the belief restrictions: unlike Moulin's (1984) complete information case, the concavity and LSE conditions alone do not suffice for the uniqueness result. A case in point is provided by Section IVB, in which we show that mechanisms that satisfy both the concavity and LSE conditions, but fail  $\mathcal{B}$ -incentive compatibility, may have multiple  $\mathcal{B}$ -rationalizable outcomes.

<sup>9</sup>In Ollár and Penta (2016), we provide a full characterization of  $\mathcal{B}$ -implementation. The characterization result, however, is not particularly suited to providing insights on the design of transfers for full implementation. Thus, rather than discussing the full characterization, we focus here on sufficient conditions which provide a clearer economic intuition. None of the results in this paper rely on the characterization in Ollár and Penta (2016).

### III. Designing Transfers for Full Implementation

Theorem 1 suggests that mechanisms with concave payoff functions and *small* strategic externalities may be useful to attain full implementation. In the following we exploit this insight to explicitly design transfers for full implementation.

We begin by considering belief-free implementation, which ensures the maximum level of robustness. In Section IIIA we introduce the *canonical transfers* and show that they characterize the mechanisms that achieve belief-free implementation. Hence, if the canonical transfers induce overly strong strategic externalities, belief-free implementation is impossible. Full implementation may still be possible if information about beliefs is used. In Section IIIB, we obtain transfers for full implementation adding a belief-based term to the canonical transfers. The extra term is derived from moment conditions chosen in order to ensure the concavity and the  $\mathcal{B}$ -LSE conditions. Full implementation then follows from Theorem 1.

#### A. Canonical Transfers and Belief-Free Implementation

Consider the following transfers: for each  $i \in I$  and  $m \in \Theta$ , let

$$(8) \quad t_i^*(m) = -v_i(d(m), m) + \int_{\underline{\theta}_i}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s_i, m_{-i}), s_i, m_{-i}) ds_i.$$

We will refer to  $t^* = (t_i^*(\cdot))_{i \in I}$  as the *canonical transfers*, and to the pair  $(d, t^*)$  as the *canonical direct mechanism*. In the canonical direct mechanism, agents pay their valuation as entailed by the reports profile minus the *total* own preference effect. This way, agents' payments coincide with the *total allocation effect* of their report, given the opponents' messages.<sup>10</sup>

The canonical transfers generalize several known mechanisms, such as the VCG mechanism if  $d$  is the efficient allocation rule; the mechanisms of Myerson (1981), Laffont and Maskin (1980) and Mookherjee and Reichelstein (1992) in private value settings; and Li (2016) and Roughgarden and Talgam-Cohen (2013) with interdependent values. Proposition 1 below shows that the canonical transfers characterize the direct mechanisms that achieve belief-free full implementation. This result follows immediately from the following lemma, which extends analogous results for partial implementation in the special cases mentioned above.

<sup>10</sup>Consider the first term of (8). Let  $\varpi_i(\theta) \equiv v_i(d(\theta), \theta)$  and consider its derivative with respect to  $\theta_i$  at  $\tilde{\theta}$ ,

$$\frac{\partial \varpi_i(\tilde{\theta})}{\partial \theta_i} = \frac{\partial v_i}{\partial x}(d(\tilde{\theta}), \tilde{\theta}) \cdot \frac{\partial d(\tilde{\theta})}{\partial \theta_i} + \frac{\partial v_i}{\partial \theta_i}(d(\tilde{\theta}), \tilde{\theta}).$$

The first term represents the *allocation effect*: the variation of  $i$ 's valuation at  $\tilde{\theta}$ , when the allocation changes due to a change in the reported type. The second term is the *own preference effect*: the variation of  $i$ 's valuation due to  $\theta_i$ , holding  $d(\tilde{\theta})$  constant. Integrating both terms with respect to  $\theta_i$ , we obtain that  $\varpi_i$  can be decomposed as

$$\varpi_i(\tilde{\theta}) = \int_{\underline{\theta}_i}^{\tilde{\theta}_i} \frac{\partial v_i}{\partial x}(d(s_i, \tilde{\theta}_{-i}), s_i, \tilde{\theta}_{-i}) \cdot \frac{\partial d(s_i, \tilde{\theta}_{-i})}{\partial \theta_i} ds_i + \int_{\underline{\theta}_i}^{\tilde{\theta}_i} \frac{\partial v_i}{\partial \theta_i}(d(s_i, \tilde{\theta}_{-i}), s_i, \tilde{\theta}_{-i}) ds_i,$$

where the first term is the *total allocation effect* and the second is the *total preference effect*. Thus, the canonical transfer in (8) can be seen as the negative of the total allocation effect of the reported type, given opponents' reports.



LEMMA 1: Suppose that  $(d, t)$  is ex post incentive compatible and differentiable. Then, for every  $i$  and for every  $m$ , there exists a function  $\tau_i : \Theta_{-i} \rightarrow \mathbb{R}$  such that  $t_i(m) = t_i^*(m) + \tau_i(m_{-i})$ .

PROPOSITION 1: Allocation rule  $d$  is belief-free fully implementable by a differentiable direct mechanism if and only if it is belief-free fully implemented by the canonical direct mechanism.

In many environments of economic interest the canonical direct mechanism is strictly concave (see Section IV). Hence, if in such environments strict ex post incentive compatibility is possible, full implementation can only fail if the canonical direct mechanism induces overly strong strategic externalities. We provide next a measure of such strategic externalities. For any  $i \in I$ , let  $W_i : M \times \Theta \rightarrow \mathbb{R}$  be such that

$$W_i(m; \theta) := \left( \frac{\partial v_i}{\partial x}(d(m), \theta) - \frac{\partial v_i}{\partial x}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m).$$

For every  $i \in I$ , define the *externality gap* as

$$(9) \quad EG_i := \max_{\theta, m, m_i} \left( \sum_{j \neq i} \left| \frac{\partial W_i}{\partial m_j}(m; \theta) \right| - \left| \frac{\partial W_i}{\partial m_i}(m_i, m_{-i}; \theta) \right| \right).$$

COROLLARY 1: Suppose that the canonical direct mechanism is strictly concave in the sense above. Then, if  $(d, t^*)$  is strictly EPIC but not belief-free fully implementable, then  $EG_i > 0$  for some  $i$ .

To understand this result, note that  $W_i(m; \theta)$  is the derivative of the ex post payoff function of the canonical direct mechanism with respect to  $i$ 's type, evaluated at state  $\theta$ , when the reported profile is  $m$ . The externality gap therefore measures the maximal difference between the opponents' ability to jointly affect this derivative and agent  $i$ 's own effect, evaluated across all possible combinations of states and reports. Hence,  $EG_i < 0$  means that  $i$ 's own effect on the first-order condition of the canonical direct mechanism always dominates the combined strategic externalities at all states and reports. The result then follows from Theorem 1.

### B. Full Implementation via Moment Conditions

By the results in Section IIIA, if the canonical direct mechanism is strictly concave and strictly ex post incentive compatible, failure to achieve belief-free implementation is due to the existence of positive externality gaps. In these cases, information about beliefs may be useful to weaken the strategic externalities and achieve full implementation. In general, also incentive compatibility may be problematic. In that case, belief restrictions can be used to ensure both properties.

The next theorem relates the possibility of achieving full implementation to the moment conditions consistent with  $\mathcal{B}$ . As discussed in Section I, the choice of the moment condition affects both the design and the degree of robustness achieved by

the mechanism. This result thus formalizes the idea that robustness in our model is envisioned as a choice of the designer.

**THEOREM 2:** Allocation rule  $d : \Theta \rightarrow X$  is fully  $\mathcal{B}$ -implementable if there exists a  $\mathcal{B}$ -consistent moment condition  $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B})$  such that, for all  $i$ ,  $\theta_i, m_i, m'_i$ , and for all  $\mu \in C_{\theta_i}^{\mathcal{B}^\rho}$ :

- (i)  $\int_{M_{-i} \times \Theta_{-i}} \frac{\partial W_i}{\partial m_i}(m_i, m_{-i}; \theta_i, \theta_{-i}) d\mu < f'_i(m_i)$ , and
- (ii)  $\int_{M_{-i} \times \Theta_{-i}} \left| \frac{\partial W_i}{\partial m_i}(m'_i, m_{-i}; \theta_i, \theta_{-i}) - f'_i(m'_i) \right| d\mu$   
 $> \sum_{j \neq i} \int_{M_{-i} \times \Theta_{-i}} \left| \frac{\partial W_i}{\partial m_j}(m_i, m_{-i}; \theta_i, \theta_{-i}) + \frac{\partial L_i}{\partial m_j}(m_{-i}) \right| d\mu.$

Moreover, for  $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B})$  that satisfies the two conditions, the following transfers guarantee full  $\mathcal{B}^\rho$ -implementation (hence full  $\mathcal{B}$ -implementation):

$$(10) \quad t_i^\rho(m) = \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{L_i(m_{-i}) m_i - \int_{\theta_i}^{m_i} f_i(s_i) ds_i}_{\text{moment condition-based term}}.$$

The following (stronger) version of these conditions is often easier to check in applications.

**Remark 2:** The conditions of Theorem 2 are satisfied if for all  $i$ , for all  $\theta \in \Theta$ , for all  $m_{-i} \in M_{-i}$ , and for all  $m_i, m'_i \in M_i$ :

- (i)  $\frac{\partial W_i}{\partial m_i}(m_i, m_{-i}; \theta) < f'_i(m_i)$ ;
- (ii)  $\left| \frac{\partial W_i}{\partial m_i}(m'_i, m_{-i}; \theta) - f'_i(m'_i) \right| > \sum_{j \neq i} \left| \frac{\partial W_i}{\partial m_j}(m_i, m_{-i}; \theta) + \frac{\partial L_i}{\partial m_j}(m_{-i}) \right|.$

Theorem 2 states two properties of moment conditions that are useful to achieve full implementation, and may thus guide the designer's choice of a suitable moment condition. To understand what these are, let us consider the ex post versions stated in Remark 2. First, note that if  $f_i$  is constant, then Condition 1 ensures that the canonical direct mechanism is strictly concave in own action and strictly EPIC. Second, note that if the externality gap (9) is negative for all  $i$ , then Condition 2 is satisfied by any trivial moment condition, in which  $f_i$  and  $L_i$  are constant functions. Since such trivial moment conditions are consistent with any belief restrictions, full implementation is guaranteed by the canonical direct mechanism in the belief-free sense. Now, suppose that the externality gap is positive for some agent. Condition 2 clarifies which properties of beliefs can be used to weaken the strategic externalities: a moment condition in which the derivative of  $f_i$  has the opposite sign of  $\partial W_i / \partial m_i$  can be used to increase the *own effect*, whereas the *external effects* can be weakened by moment functions  $L_i$  with derivatives that contrast the strategic externality in the canonical direct mechanism. Condition 1 instead requires that the own effect in the canonical direct mechanism is bounded above by the derivative of  $f_i$ .

To gain further insights on how these conditions contribute to the full implementation result, it is useful to consider the transfers that achieve full implementation

(equation (10)). With these transfers, the first-order derivative of  $\theta_i$ 's expected payoff, given  $\mu \in \Delta(M_{-i} \times \Theta_{-i})$ , is

$$\frac{\partial EU_{\theta_i}^\mu(m_i)}{\partial m_i} = \int_{M_{-i} \times \Theta_{-i}} \left[ \left( \frac{\partial v_i}{\partial x}(d(m), \theta) - \frac{\partial v_i}{\partial x}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) + L_i(m_{-i}) - f_i(m_i) \right] d\mu.$$

This shows that for any conjectures  $\mu \in C_{\theta_i}^{\mathcal{B}^\rho}$  concentrated on the opponents' truthful profile, the report  $m_i = \theta_i$  satisfies the first-order conditions. This does not necessarily result in strict  $\mathcal{B}^\rho$ -IC, as that also depends on the second-order conditions. But Condition 1 guarantees that the ensuing mechanism is concave, and hence the second-order conditions are met. This mechanism therefore is strictly  $\mathcal{B}^\rho$ -IC and satisfies the concavity condition in Theorem 1. Full implementation follows from the fact that Condition 2 in Theorem 2 also guarantees the  $\mathcal{B}^\rho$ -LSE condition of Theorem 1.

**Example 7 (Example 1 Redux):** Note that applying the formula of the canonical transfers (8) to Example 1, and dropping all terms that are constant in  $i$ 's own report (and hence do not affect his best response), delivers the VCG transfers in (1). It is easy to verify that, in this case,  $\frac{\partial W_i}{\partial m_i}(m; \theta) = -(1 + \gamma)$  and  $\frac{\partial W_i}{\partial m_j}(m; \theta) = -(1 + \gamma)\gamma$ . Letting  $L_i(\theta_j) = (1 + \gamma)\gamma\theta_j$ , under the independent uniform common prior we obtain  $f_i(\theta_i) := E(L_i(\theta_i) | \theta_i) = 0.5 \cdot (1 + \gamma)\gamma$ , and hence  $\frac{\partial L_i}{\partial \theta_j}(\theta_j) = (1 + \gamma)\gamma$  and  $f_i'(\theta_i) = 0$ . Thus, both conditions of Theorem 2 are satisfied for  $\rho = (L_i, f_i)_{i \in \{1, 2\}}$ , and in fact with the RHS of Condition 2 equal to 0. Clearly, these moment conditions only rely on common knowledge that  $E(\theta_j | \theta_i) = 0.5$ , and applying the formula in (10) to moment condition  $\rho = (L_i, f_i)_{i \in \{1, 2\}}$  we obtain the adjusted transfers (3) in Example 1.

#### IV. Applications and Extensions

In this section we illustrate how Theorem 2 can be applied to special cases of interest, under different assumptions on agents' beliefs. We also show further robustness properties of the design strategy put forward in Theorem 2.

##### A. SCC-Environments: A Robustness Trade-Off

For simplicity, in this subsection we maintain that  $X \subseteq \mathbb{R}$ . A common assumption in applications is the following *single-crossing condition* (SCC).

ASSUMPTION 1 (SCC): For all  $i \in I$  and  $(x, \theta)$ ,  $\frac{\partial^2 v_i}{\partial x \partial \theta_i}(x, \theta) > 0$ .

The next lemma generalizes standard results on ex post (partial) implementation.

LEMMA 2: In environments that satisfy Assumption 1, the canonical direct mechanism  $(d, t^*)$  is strictly ex post incentive compatible if and only if  $d$  is strictly increasing in each  $\theta_i$ .

Because of this result, in the following we refer to *SCC-environments* as those that (in addition to the maintained assumptions) satisfy Assumption 1 and such that  $d$  is *strictly increasing* in each  $\theta_i$ . The next result, which follows immediately from Lemma 2 and Corollary 1, summarizes easy-to-check conditions for belief-free *full* implementation in SCC-environments.<sup>11</sup>

**PROPOSITION 2:** *In SCC-environments, the allocation rule  $d$  is belief-free fully implementable if  $\frac{\partial W_i}{\partial m_i}(m; \theta) < 0$  and  $\left| \frac{\partial W_i}{\partial m_i}(m; \theta) \right| > \sum_{j \neq i} \left| \frac{\partial W_i}{\partial m_j}(m'_i, m_{-i}; \theta) \right|$  for all  $i, \theta, m$ , and  $m'_i$ .*

Hence, in SCC-environments, belief-free full implementation may fail only if the canonical direct mechanism is not globally concave or if there are positive externality gaps. Proposition 2 therefore highlights a trade-off in SCC-environments, between the robustness of the partial implementation result—obtained by the canonical direct mechanism in a belief-free sense—and the possibility of achieving full implementation: the latter necessarily relies on belief restrictions and therefore reduces the robustness of the partial implementation result.

To simplify the analysis, we first consider the following assumptions on the valuation functions  $v_i$ , and the allocation rule  $d$ .

**ASSUMPTION 2 (SCC-PC):** (i) *For each  $i$  and  $j$ ,  $\partial^2 v_i / \partial^2 x$  and  $\partial^2 v_i / \partial x \partial \theta_j$  are constant in  $\theta$ ; (ii)  $d$  is linear in  $\theta$ :  $\frac{\partial^2 d}{\partial \theta_i \partial \theta_j}(\theta) = 0$  for all  $i, j$ , and  $\theta$ .*

A special case of these conditions is provided by environments with quadratic valuations and linear allocation functions.<sup>12</sup> Assumption 2 also accommodates more general dependence on  $x$ , as long as the concavity and the cross derivatives are public information. We thus refer to the SCC-environments which also satisfy Assumption 2 as *SCC-environments with public concavity* (SCC-PC). This assumption, however, is not essential to our analysis. In Section IVA we discuss how the results are affected when it is relaxed. Note that, in SCC-PC environments,  $\frac{\partial W_i}{\partial m_i}(m; \theta) < 0$  for all  $(m, \theta)$ . Hence, by Proposition 2, belief-free implementation fails only if there are positive externality gaps; in this case if  $\left| \frac{\partial^2 v_i}{\partial x \partial \theta_i}(d(m), m) \right| \leq \sum_{j \neq i} \left| \frac{\partial^2 v_i}{\partial x \partial \theta_j}(d(m), m) \right|$  for some  $m$  and  $i$ .

<sup>11</sup> This result is related to Bergemann and Morris (2009a)—henceforth, BM—who characterize belief-free implementation via direct mechanisms in environments with monotone aggregators (i.e., such that  $\forall i, v_i(x, \theta) = w_i(x, h_i(\theta))$  for some  $w_i : X \times \mathbb{R} \rightarrow \mathbb{R}$  and  $h_i : \Theta \rightarrow \mathbb{R}$  strictly increasing in  $\theta_i$ ). BM's characterization is in terms of strict EPIC and the following contraction property (BM, Definition 5):  $\forall \beta : \Theta \rightarrow 2^\Theta$  such that  $\theta \in \beta(\theta)$  for all  $\theta$ , but  $\beta(\theta') \neq \{\theta'\}$  for some  $\theta'$ , there exists  $i, \theta_i$ , and  $\theta'_i \in \beta_i(\theta_i)$  with  $\theta'_i \neq \theta_i$  such that, for all  $\theta_{-i}$  and  $\theta'_{-i} \in \beta_{-i}(\theta_{-i})$ ,  $\text{sign}(\theta_i - \theta'_i) = \text{sign}(h_i(\theta_i, \theta_{-i}) - h_i(\theta'_i, \theta'_{-i}))$ . In SCC-environments that satisfy both BM's aggregator property (which we do not assume) and our smoothness assumptions (Section I), it can be shown that our condition implies BM's contraction property. The converse holds if the aggregators are also *symmetric* (i.e., such that  $\partial h_i(\theta) / \partial \theta_i = \partial h_j(\theta) / \partial \theta_j$  and  $\sum_{k \neq i} (\partial h_i(\theta) / \partial \theta_k) = \sum_{k \neq j} (\partial h_j(\theta) / \partial \theta_k)$  for all  $i, j$ , and  $\theta$ ) and the environment satisfies our Assumption 2. Environments with symmetric aggregators include the examples in Section IA, where our condition also coincides with Chung and Ely's (2001) sufficient condition for environments with linear aggregators.

<sup>12</sup> Quadratic-linear models are common in the literature, as they ensure linear best-responses. Examples include social interactions models (e.g., Blume et al. 2015), markets with network externalities (e.g., Fainmesser and Galeotti 2015), supply function competition (e.g., Vives 2011), divisible good auctions (e.g., Wilson 1979) and public goods (e.g., Duggan and Roberts 2002).

*Common Prior Models.*—As explained in Section I, in common prior environments the belief restrictions  $\mathcal{B}$  are such that for every  $i$  and  $\theta_i$ ,  $B_{\theta_i} = \{b_{\theta_i}\}$ , where  $b_{\theta_i} = p(\cdot | \theta_i) \in \Delta(\Theta_{-i})$  for some  $p \in \Delta(\Theta)$ . Then,  $\mathcal{B}$ -IC coincides with interim-IC, and one could similarly refer to  $\mathcal{B}$ -DS implementation as *interim* dominant strategy implementation. (If  $\mathcal{B}$  is a standard common prior model,  $\mathcal{B}$ -DS implementation is equivalent to truthful revelation being strictly dominant in the interim normal form of the Bayesian game.) Since, in the following, the common prior assumption is maintained, we will take expectations without making the prior explicit. So, for instance, given a function  $L_i : \Theta_{-i} \rightarrow \mathbb{R}$ , we will simply write  $E(L_i(\theta_{-i}) | \theta_i)$  instead of  $\int_{\Theta_{-i}} L_i(\theta_{-i}) dp(\cdot | \theta_i)$ .

*Independent Types.*—In an independent common prior model, for any  $L_i : \Theta_{-i} \rightarrow \mathbb{R}$ , the condition  $E(L_i(\theta_{-i}) | \theta_i) = f_i(\theta_i)$  holds true with  $f_i : \Theta_i \rightarrow \mathbb{R}$  such that  $f'_i = 0$  (by the definition of independence). Hence, since  $L_i$  can be chosen freely in common prior models, independence leaves us enough leeway to manipulate the external effects on the RHS of Condition 2 of Theorem 2, without affecting the LHS. In particular, the ex post condition of Remark 2 can be rewritten as

$$(11) \quad \left| \left( \frac{\partial^2 v_i}{\partial x \partial \theta_i}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) \right| > \sum_{j \neq i} \left| \left( -\frac{\partial^2 v_i}{\partial x \partial \theta_j}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) + \frac{\partial L_i}{\partial m_j}(m_{-i}) \right|.$$

Hence, in this case we can completely neutralize the strategic externalities, setting the RHS of this inequality equal to zero, by choosing  $\hat{L}_i$  such that

$$(12) \quad \frac{\partial \hat{L}_i}{\partial m_j}(m_{-i}) = \left( \frac{\partial^2 v_i}{\partial x \partial \theta_j}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) \text{ for all } m \text{ and } j \neq i, \text{ or}$$

$$(13) \quad \hat{L}_i(m_{-i}) = \sum_{j \neq i} \left( \frac{\partial^2 v_i}{\partial x \partial \theta_j}(d(m), m) \cdot m_j \right) \frac{\partial d}{\partial \theta_i}(m).$$

(Equations (12) and (13) are well defined because Assumption 2 ensures that the RHS of (12) is constant in  $m_i$ .) Hence, the following proposition holds.

**PROPOSITION 3:** *In SCC-PC-environments with an independent common prior, the following transfers ensure full implementation in interim dominant strategies:*

$$(14) \quad t_i^\rho(m) = \underbrace{t_i^*(m)}_{\text{canonical transfers}} + \underbrace{\hat{L}_i(m_{-i}) m_i}_{\substack{\text{compensation for} \\ \text{strategic externality} \\ \text{(depends on } m_i \text{ and } m_{-i})}} - \underbrace{\int_{\theta_i}^{m_i} E(\hat{L}_i(\theta_{-i}) | s_i) ds_i}_{\substack{\text{belief-based adjustment: cumulative} \\ \text{expected marginal compensation} \\ \text{(only depends on } m_i)}}.$$

To understand the logic of the mechanism, consider the function  $\hat{L}_i(m_{-i})$  defined in (13). The term in parentheses represents the effect of  $j$ 's report on  $i$ 's marginal utility for  $x$ , and is multiplied by the impact of  $i$ 's report on the allocation. Overall, this is the total strategic externality that agent  $i$  is subject to, for each increment of his own report. The transfers in (14) therefore are such that, starting from the

canonical direct mechanism, agent  $i$  is compensated for the total strategic externality that other players impose on him. The last term in (14) is the expected marginal effect on such a compensation, when  $i$  reports  $m_i$ . This term is added to prevent the agent from misreporting his type, in order to inflate the compensation for the strategic externality. Hence, the first term eliminates the strategic externalities, and the second restores incentive compatibility.

*Affiliated Types.*—Under the maintained assumptions for SCC-PC-environments, and if valuations are *supermodular* (that is,  $0 < \frac{\partial^2 v_i}{\partial x \partial \theta_j}(x, \theta) < \infty$  for all  $i, j, x$ , and  $\theta$ ), the moment functions  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$  defined in (13) are strictly increasing in  $m_j$  for each  $j \neq i$ . Then, if types are *affiliated*, Theorem 5 in Milgrom and Weber (1982) implies that  $E(\hat{L}_i(\theta_{-i}) | \theta_i)$  is increasing in  $\theta_i$ . Hence, letting  $\hat{f}_i(\cdot) \equiv E(\hat{L}_i(\theta_{-i}) | \cdot)$ , the moment condition  $\rho = (\hat{L}_i, \hat{f}_i) \in \varrho(\mathcal{B})$  satisfies  $\hat{f}'_i > 0$  for all  $i$ . By construction,  $\hat{L}_i$  is such that the RHS of Condition 2 in Theorem 2 is equal to zero. Since  $\hat{f}'_i > 0$ , SCC implies that the LHS of the same condition is positive. The next result follows.

**PROPOSITION 4:** *In SCC-PC-environments with affiliated types and supermodular valuations, the transfers in (14) ensure full implementation in interim dominant strategies.*

*Equivalence of EPIC and DS-implementability.*—The results above can also be used to derive an equivalence between ex post incentive compatibility and dominant strategy implementation.

**PROPOSITION 5:** *In independent common prior environments that satisfy Assumptions 1 and 2, an allocation function is interim dominant strategy implementable if and only if the canonical direct mechanism is strictly ex post incentive compatible. If valuations are supermodular, the equivalence extends to affiliated types.*

In the proof, in the Appendix, first we show that an allocation rule is interim dominant strategy implementable only if it is strictly increasing. The “only if” part then follows from Lemma 2. Conversely, if  $(d, t^*)$  is strictly EPIC, then  $d$  is strictly increasing by Lemma 2. Propositions 3 and 4 in turn imply that the allocation rule is interim dominant strategy implementable.

Proposition 5 is related to results by Manelli and Vincent (2010)—henceforth, MV—and Gershkov et al. (2013) which show that, in Bayesian environments with *private values*, for any interim incentive-compatible mechanism there is an *equivalent* mechanism that is dominant strategy incentive compatible. Given the restriction to private values, those results can be interpreted as an equivalence between *partial* and *full* implementation in direct mechanisms. From this viewpoint, Proposition 5 can be seen as a generalization of that insight to Bayesian environments with *interdependent values*.<sup>13</sup> MV’s notion of equivalence, however, is different from ours. In

<sup>13</sup>We are grateful to Stephen Morris for this insight.



particular, MV define two mechanisms as equivalent if they deliver the same interim expected utilities for all agents and the same ex ante expected social surplus. Here instead we maintain the traditional notion of equivalence, which requires that the mechanisms induce the same ex post allocation. (As shown by Gershkov et al. 2013, equivalence results à la MV do not extend beyond environments with linear utilities and independent types.)

*Moment Conditions without a Prior.*—In real-world problems of mechanism design, the designer's information typically does not take the form of a common prior distribution on agents' types. For instance, when the designer's information is based on econometric estimates, and if such estimates are assumed common knowledge, then the belief restrictions  $\mathcal{B}$  are naturally represented directly in terms of a set of moment conditions (cf. Section IB). In these cases, it may be interesting to ask which moments it would be useful to estimate, provided they are common knowledge.<sup>14</sup> The next result shows that, in SCC-PC-environments, the conditional expectations  $E(\theta_j | \theta_i)$  are all that is needed for the implementation result.

**PROPOSITION 6:** *Consider a SCC-PC-environment with supermodular valuations. Let the belief restrictions  $\bar{\mathcal{B}}$  be such that only common knowledge of the conditional expectations  $E(\theta_j | \theta_i)$  is maintained, for all  $i$  and  $j$ . If such conditional expectations are differentiable and nondecreasing in  $\theta_i$  for each  $i$  and  $j$ , then the transfers in equation (14) ensure  $\bar{\mathcal{B}}$ -implementation in  $\bar{\mathcal{B}}$ -dominant strategies.*

The proof, in the Appendix, is based on the observation that the function  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$  defined in equation (13) is linear if the environment satisfies the SCC-PC conditions. Hence, if the conditional expectations  $E(\theta_j | \theta_i)$  are common knowledge in  $\bar{\mathcal{B}}$ , so are the conditional expectations  $E(\hat{L}_i(\theta_{-i}) | \theta_i)$ , which can thus be used as moment conditions to weaken the strategic externalities, similar to what we did for the common prior environments in Propositions 3 and 4.

*Discussion.*—The logic of Propositions 3, 4, and 5 extends beyond the cases of common prior models with independent or affiliated types. To see this, notice that for  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$  defined in equation (13), the maintained assumptions for SCC-PC-environments guarantee that the RHS of Condition 2 in Theorem 2 is equal to 0. Affiliation or independence further guarantee that the conditional moment  $E(\hat{L}_i(\theta_{-i}) | \theta_i)$  is nondecreasing in  $\theta_i$ . Hence, letting  $\hat{f}_i(\theta_i) := E(\hat{L}_i(\theta_{-i}) | \theta_i)$ , the moment condition  $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$  can be used with no risk of upsetting the LHS of Condition 2 in Theorem 2. This argument, however, remains valid whenever  $E(\hat{L}_i(\theta_{-i}) | m_i) < \frac{\partial W_i}{\partial m_i}(m; \theta)$  for all  $m$ , which ensures that both conditions of Theorem 2 are satisfied by  $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$ . In Proposition 6, the assumption that  $E(\theta_j | \theta_i)$  are nondecreasing in  $\theta_i$  plays the same role as the assumptions of independence and affiliation in the common prior models, and can be weakened similarly.

<sup>14</sup> Ollár and Penta (2017) consider a different non-Bayesian setting, in which types are commonly believed to follow the same distribution, but the distribution itself is unknown.

Assumption 2 may also be weakened in Propositions 3, 4, and 6. In the argument above, we used the assumption to ensure  $\partial W_i / \partial m_i < 0$  and that  $\hat{L}_i$  could be designed to completely neutralize the strategic externalities of the canonical direct mechanism. Clearly,  $\partial W_i / \partial m_i < 0$  can be guaranteed by weaker conditions. If Assumption 2 is violated, however, then we may not be able to completely offset the strategic externalities. But if  $|\partial^2 d / \partial \theta_i \partial \theta_j|$  and the variations of the valuations' concavity are small relative to  $|\partial W_i / \partial m_i|$ , then  $\hat{L}_i$  can still be chosen so that the RHS of (11) is bounded above by  $|\partial W_i / \partial m_i|$ , and the argument remains valid. The only difference is that full implementation would not occur in one round of  $\mathcal{B}$ -rationalizability: that is, the transfers would ensure  $\mathcal{B}$ -implementation, but not in  $\mathcal{B}$ -dominant strategies.

### B. Sensitivity Analysis and Approximate Moment Conditions

The implementation result in Theorem 2 is *robust* in the sense that only common knowledge of a certain moment condition is required. But what if such a moment condition is not exactly satisfied? How sensitive are the implementation results to possible misspecifications of the moment condition? In this section we show that our design strategy also ensures that the mechanism is well behaved with respect to small misspecifications of the moment conditions.

**Example 8 (Sensitivity Analysis):** Consider the environment in Example 2. Adopting the design strategy of Theorem 2, we showed that the strategic externalities of the VCG mechanism could be sufficiently reduced, so as to induce contractive best responses, adopting the moment condition  $\gamma \kappa E(\theta_k - \theta_j | \theta_i) = 0$ . But what if  $\gamma \kappa E(\theta_k - \theta_j | \theta_i)$  is only within  $\varepsilon$  of 0, so that the moment condition is not exactly satisfied? Then, for any  $\theta_i$ , the set of rationalizable reports consistent with common belief that  $\gamma \kappa E(\theta_k - \theta_j | \theta_i) \in [0 \pm \varepsilon] := [-\varepsilon, +\varepsilon]$  is equal to  $R_i^{\mathcal{B}}(\theta_i) = \left[ \theta_i \pm \frac{1}{(1 - |\gamma + \delta|) \kappa} \cdot \varepsilon \right]$ .

Thus, small misspecifications of the moment condition induce small misreports, and hence (given the continuity of  $d$ ) small misallocation relative to the designer's objective. Moreover, the impact of the misspecified moment condition is increasing in  $|\gamma + \delta|$ , which measures the strategic externalities in the belief-based mechanism (and such that  $|\gamma + \delta| < 1$  in Example 2), and decreasing in the concavity in own-action, captured by  $\kappa \equiv (1 + \gamma + \delta) > 0$ . Thus, the resilience to such misspecifications is improved by mechanisms with smaller strategic externalities, and maximally so if they achieve  $\mathcal{B}$ -DS implementation, or by mechanisms with larger concavity in own action. (In Example 1, the rationalizable reports if the moment condition is misspecified are  $R_i^{\mathcal{B}}(\theta_i) = [\theta_i \pm \varepsilon / (1 + \gamma)]$ .)

Our next result generalizes these insights. In particular, we show that concave mechanisms with limited strategic externalities ensure continuity with respect to misspecifications of the moment conditions, and we characterize the impact of such misspecifications, relating it to the strength of the strategic externalities. To this end, consider the transfers in (10), and suppose that  $\rho = (L_i, f_i)_{i \in I}$  satisfies the conditions

of Theorem 2. Then, for any  $\theta_i \in \Theta_i$ , we define the smallest own-concavity and strongest strategic externality for  $\theta_i$ , respectively as

$$OC_i^\rho(\theta_i) := \min_{(m'_i, \mu) \in M_i \times C_{\theta_i}^{B^\rho}} \int_{M_{-i} \times \Theta_{-i}} \left| \frac{\partial W_i}{\partial m_i}(m'_i, m_{-i}; \theta_i, \theta_{-i}) - f'_i(m'_i) \right| d\mu \quad \text{and}$$

$$SE_i^\rho(\theta_i) := \max_{(m'_i, \mu) \in M_i \times C_{\theta_i}^{B^\rho}} \sum_{j \neq i} \int_{M_{-i} \times \Theta_{-i}} \left| \frac{\partial W_i}{\partial m_j}(m'_i, m_{-i}; \theta_i, \theta_{-i}) + \frac{\partial L_i}{\partial m_j}(m_{-i}) \right| d\mu.$$

We also define the *overall own-concavity* and *normalized strategic externalities*, respectively, as  $OC^\rho := \min_{i \in I, \theta_i \in \Theta_i} OC_i^\rho(\theta_i)$ , and  $NSE^\rho = \max_{i \in I, \theta_i \in \Theta_i} \frac{SE_i^\rho(\theta_i)}{OC_i^\rho(\theta_i)}$ . (In Examples 2 and 8,  $OC^\rho = \kappa$  and  $NSE^\rho = |\delta + \gamma| < 1$ . In Examples 1 and 7,  $NSE^\rho = 0$  and  $OC^\rho = |1 + \gamma|$ .) Notice that  $OC_i^\rho(\theta_i)$  and  $SE_i^\rho(\theta_i)$  correspond, respectively, to the LHS and RHS of Condition 2 in Theorem 2. Hence, under the conditions of Theorem 2,  $OC^\rho > 0$  and  $NSE^\rho \in [0, 1)$ . In this case, best responses are contractive, and more so as  $NSE^\rho$  gets smaller. The difference  $(1 - NSE^\rho)$  therefore provides a measure of contractiveness. We next show that these terms also capture the sensitivity to misspecifications of the moment condition  $\rho$ .

**THEOREM 3:** Suppose that  $\rho = (L_i, f_i)_{i \in I}$  satisfies Conditions (i) and (ii) of Theorem 2, but it is only approximately satisfied in  $\mathcal{B}$ : that is, for all  $i, \theta_i$ , and  $b_i \in B_{\theta_i}$ ,  $E_{b_i}(L_i(\theta_{-i})) \in [f_i(\theta_i) \pm \varepsilon]$  for some  $\varepsilon > 0$ . Then, the transfers  $(t_i^\rho)_{i \in I}$  defined in (10) achieve almost truthful  $\mathcal{B}$ -implementation. That is, for all  $i$  and  $\theta_i \in \Theta_i$ ,  $R_i^{\mathcal{B}}(\theta_i) \subseteq \left[ \theta_i \pm \frac{1}{(1 - NSE^\rho) OC^\rho} \varepsilon \right]$ .

This result implies a convenient continuity property: as the misspecification of the moment condition vanishes ( $\varepsilon \rightarrow 0$ ), the mechanism approaches truthful implementation. Moreover, for given  $\varepsilon > 0$ , deviations from truthful implementation decrease with the own-concavity and increase with the strategic externalities. The latter effect is minimal when  $NSE^\rho = 0$ , that is if the mechanism achieves dominant-strategy implementation. Hence, Theorem 3 provides further reasons for pursuing the design of concave mechanisms with limited strategic externalities.

Theorem 3 can also be seen as a generalization of Theorem 2 to accommodate *approximate* moment conditions. Formally, for any  $\mathcal{B}$  and  $\varepsilon \geq 0$ , let  $\varrho(\mathcal{B}, \varepsilon)$  denote the set of moment conditions that are approximately consistent with  $\mathcal{B}$ :  $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B}, \varepsilon)$  if and only if for all  $i, \theta_i$  and  $b_i \in B_{\theta_i}$ ,  $E_{b_i}(L_i(\theta_{-i})) \in [f_i(\theta_i) \pm \varepsilon]$ . Clearly,  $\varrho(\mathcal{B}, 0) = \varrho(\mathcal{B})$  and the set  $\varrho(\mathcal{B}, \varepsilon)$  increases with  $\varepsilon$ : for any  $\mathcal{B}$ ,  $\varepsilon' > \varepsilon$  implies  $\varrho(\mathcal{B}, \varepsilon) \subseteq \varrho(\mathcal{B}, \varepsilon')$ . Then the following result holds.

**COROLLARY 2:** Let  $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B}, \varepsilon)$  satisfy Conditions (i) and (ii) of Theorem 2. Then, the transfers  $(t_i^\rho)_{i \in I}$  defined in (10) ensure that for all  $i$  and  $\theta_i \in \Theta_i$ ,  $R_i^{\mathcal{B}}(\theta_i) \subseteq \left[ \theta_i \pm \frac{1}{(1 - NSE^\rho) OC^\rho} \varepsilon \right]$ .

Hence, for *exact* moment conditions ( $\varepsilon = 0$ ), we obtain the truthful implementation result of Theorem 2 as a special case. As  $\varepsilon$  increases and approximate moment conditions are included, misreports are possible, but they remain small and thus

(given the continuity of  $d$ ) ensure that the allocation stays close to the designer's objective. Corollary 2 thus suggests a notion of approximate implementation reminiscent of virtual implementation, but with the difference that here the allocation is guaranteed to be nearby in the allocation space, rather than the space of lotteries.

## V. Related Literature

Our work is related to several strands of the literature in game theory and mechanism design. We briefly discuss the most closely related literature.

*Solution Concept.*—As explained in Section II,  $\mathcal{B}$ -rationalizability is a special case of  $\Delta$ -rationalizability (Battigalli 2003 and Battigalli and Siniscalchi 2003), and generalizes several versions of rationalizability for incomplete information games, including Bergemann and Morris's (2009a) belief-free version (obtained letting  $\mathcal{B} = \mathcal{B}^{BF}$ ) and Dekel, Fudenberg, and Morris's (2007) interim correlated rationalizability (ICR), if  $\mathcal{B}$  is a standard Bayesian model. ICR has also been studied by Weinstein and Yildiz (2007, 2011, 2013) and Penta (2012, 2013). Battigalli et al. (2011) provide a thorough analysis of the connections between these and other versions of rationalizability.

*Full Implementation.*—Within the vast literature on full implementation, the closest papers are Bergemann and Morris (2009a) and Oury and Tercieux (2012), which study implementation in belief-free rationalizability and ICR, respectively. Both belief-free and ICR-implementation are special cases of ours, with the proviso that Oury and Tercieux (2012) do not restrict attention to direct mechanisms. The restriction to direct mechanisms is also shared by Bergemann and Morris (2009a), while Bergemann and Morris (2011) study belief-free implementation in general mechanisms. Within the classical literature, Jackson (1991) and Postlewaite and Schmeidler (1986) are also connected, as our results imply Bayes-Nash implementation in Bayesian environments. From a conceptual viewpoint, our departure from that literature is inspired by Jackson's (1992) critique of unbounded mechanisms, although we push the concern for *relevance* a bit further, requiring that full implementation be achieved via transfer schemes.<sup>15</sup> In a complete information setting with quadratic preferences, Bergemann and Morris (2007) show that an ascending auction may reduce strategic uncertainty relative to its sealed-bid counterpart, thereby making full implementation easier (in the symmetric example with  $n$  agents, full implementation is possible in the ascending auction if  $\gamma < 1$ , as opposed to  $\gamma < 1/(n - 1)$  in the static auction). That insight, however, relies on the complete information assumption (see Penta 2015) and is orthogonal to the reduction of strategic externalities we pursue here (which ensures implementation for all  $\gamma$  in the examples).

*Robust Mechanism Design.*—As already discussed, most of the literature on robust mechanism design has focused on the belief-free case (see footnote 2). In

<sup>15</sup> d'Aspremont, Crémer, and Gérard-Varet (1998) also studied full implementation in environments with transferable utility, but they resort to unbounded mechanisms of the kind criticized above. Duggan and Roberts (2002) fully implement the efficient allocation of pollution via transfers, but under complete information and richer reports.

particular, Bergemann and Morris (2005, 2009a, b) study belief-free implementation in static settings, respectively in the partial, full, and virtual implementation sense. The belief-free approach has been extended to dynamic settings by Müller (2016) and Penta (2015). Penta (2015) considers dynamic mechanisms in environments in which agents may obtain information over time, and applies a dynamic version of rationalizability based on a backward induction logic (Penta 2011). Müller (2016) instead considers virtual implementation via dynamic mechanisms, in the same (static) belief-free environments as Bergemann and Morris (2009b), using a stronger version of rationalizability with forward induction. Thanks to the stronger assumptions on the belief revision policy, he shows that dynamic mechanisms weaken the conditions for virtual implementation.

Beyond the belief-free literature, Guo and Yannelis (2017) and Lopomo, Rigotti, and Shannon (2014) consider belief restrictions analogous to ours, to study, respectively, full and partial implementation, but with different notions of robustness that involve ambiguity. Artemov, Kunimoto, and Serrano (2013) also maintain some restrictions on beliefs, but focus on virtual implementation. Jehiel, Meyer-ter-Vehn, and Moldovanu (2012) show that, under certain restrictions on preferences, minimal notions of robustness are as demanding as the belief-free case when types are multidimensional, and Jehiel et al. (2006) show that ex post incentive compatibility is generically impossible for multidimensional types. This suggests that, when  $\mathcal{B}$  is not a standard type space, the one-dimensionality of  $\Theta_i$  is important for our results.

Alternative approaches to robust mechanism design have been put forward by Borgeers and Smith (2012, 2014), who show the role of eliciting beliefs to weakly implement a correspondence in a belief-free setting, or by Carroll (2015), Yamashita (2015), and Wolitzky (2016), who approach robustness from a max-min perspective. Kos and Messner (2015) also pursue the max-min approach, but in a setting in which—similar to one of our applications (Section IVA)—only the types' expected values are known.

*Mechanism Design with Transferable Utility (TU).*—TU-environments are the typical domain of the partial implementation literature. Within this area, the closest works are those that allow for interdependent values (e.g., Crémer and McLean 1985, 1988; Dasgupta and Maskin 2000; McLean and Postlewaite 2004).<sup>16</sup> In recent years, standard implementation problems have been revisited imposing extra desiderata on the mechanisms. Deb and Pai (2017), for instance, pursue symmetry of the mechanism. Mathevet (2010) and Mathevet and Taneva (2013) instead pursue supermodularity. In those papers, the extra desiderata are achieved by adding a belief-dependent component to some baseline payments, much as we attain full implementation appending an extra term to the canonical transfers.<sup>17</sup> Those papers however maintain that types are independently distributed, whereas we allow general correlations as well as weaker restrictions on beliefs. At a more technical level,

<sup>16</sup> McLean and Postlewaite (2002) explore related ideas in environments without transferable utility.

<sup>17</sup> Early examples of this principle are the mechanisms of d'Aspremont and Gérard-Varet (1979) and of Crémer and McLean (1985), which append the baseline VCG mechanism with a belief-based component in order to achieve budget balance and surplus extraction, respectively.

our design results in a contractive mechanism. Healy and Mathevet (2012) also pursue contractiveness of the mechanism, but with complete information.

## VI. Concluding Remarks

The objective of full implementation is to solve the problem of multiplicity in mechanism design. In this paper we developed an approach to full implementation which subsumes as special cases the notions of belief-free (Bergemann and Morris 2009a) and ICR-implementation (Oury and Tercieux 2012), and accommodates more realistic assumptions on agents' beliefs, intermediate between the belief-free and the classical Bayesian benchmarks. In Bayesian settings, which are standard in the applied and classical literature, our conditions also ensure Bayes-Nash implementation (e.g., Jackson 1991). The main innovation is that we achieve these results through mechanisms which are as simple as those developed by the partial implementation literature, thereby bridging two branches of the literature which have typically proceeded in parallel.

While largely inspired by the literature on belief-free mechanism design, we departed from it in many ways. The capability of our framework to accommodate general belief restrictions was key to go beyond the existing characterizations, and to provide constructive results on what can still be achieved when agents' preferences violate the conditions for belief-free implementation. The key idea is to focus on the strategic externalities rather than preferences, and to use moment conditions to induce contractive best replies. Our results suggest a clear design principle: start with the canonical transfers and then add a belief-based component to weaken the strategic externalities which may otherwise impair the full implementation result. The resulting mechanism is contractive and induces truthful revelation as the only rationalizable outcome.

As shown in Section IVB, mechanisms with small strategic externalities ensure further robustness properties, in that small misspecifications of the moment conditions result in allocations that are proportionately close to the desired one. Though beyond the scope of this paper, this suggests that the logic of our construction may be extended to moment conditions with inequalities. The notion of closeness here is in terms of the natural allocation space, as opposed to the lottery space of the virtual implementation literature, which points to a novel notion of approximate implementation which may be of independent interest for future research.

Other directions for future research include extending the analysis to nondifferentiable allocation rules, such as for instance in standard auction problems. A sensible first step in this direction might be to apply our analysis to smooth approximations of such nondifferentiable allocation rules. But the idea of using belief-restrictions to weaken the strategic externalities seems more broadly appealing, and there may be direct ways of formalizing it in the context of nonsmooth environments. The restriction to direct mechanisms also entails some loss of generality, and thus stronger results could be obtained with more general mechanisms. But we already know that some loss of generality is necessary, if we want to avoid the unrealistically complex mechanisms of the classical literature. Thus, a key challenge in pursuing this direction of research is to combine the increased generality with the ability to provide clear economic insights.



In Section V we discussed some implications of our main results for some special cases, such as environments with single-crossing preferences, with and without common priors. In common prior environments, we provided sufficient conditions for full implementation with independent and correlated types, as well as for an equivalence of partially and fully implementable allocation rules. In environments with *public concavity*, our construction indeed ensures that strategic externalities are completely eliminated, thereby achieving dominant strategy implementation. When this is the case, our results also imply max-min implementation (e.g., Carroll 2015 and Wolitzky 2016).

## APPENDIX: PROOFS

### PROOF OF THEOREM 1:

Let  $l := \max_{i, \theta_i} \left\{ \max_{m_i \in R_i^B(\theta_i)} |m_i - \theta_i| \right\}$  denote the largest distance between the truthful and some other rationalizable report, across all types. Note that  $l$  is well defined by properties of  $\Theta$  and by the maintained assumptions on  $v$ ,  $d$ , and  $t$ . By contradiction, suppose that  $l > 0$ , and let  $i$ ,  $\theta_i^*$ , and  $m_i^* \in R_i^B(\theta_i^*)$  be such that  $|m_i^* - \theta_i^*| = l$ . Since  $m_i^* \in R_i^B(\theta_i^*)$ ,  $\exists \mu \in C_{\theta_i^*}^B \cap \Delta(R_{-i}^B) : m_i^* \in \arg \max_{m_i} EU_{\theta_i^*}^\mu(m_i)$ . (This standard fixed-point property of  $\mathcal{B}$ -rationalizability follows from the maintained assumptions on  $v$ ,  $d$ , and  $t$ , which ensure that our mechanisms induce compact games with bounded and continuous payoff functions, e.g., Arieli 2010.) By  $\mathcal{B}$ -IC we also know that  $\theta_i \in R_i^B(\theta_i)$  for all  $\theta_i$  and  $i$ , hence the set of truthtelling conjectures  $C_i^T := \left\{ \mu \text{ such that } \mu\{(\theta_{-i}, m_{-i}) : m_{-i} = \theta_{-i}\} = 1 \right\} \subseteq \Delta(R_{-i}^B)$ . Let  $\mu^* \in C_i^T$  be such that  $\arg_{\Theta_{-i}} \mu^* = \arg_{\Theta_{-i}} \mu$ . Then,  $\mathcal{B}$ -IC implies that  $\theta_i^* \in \arg \max_{m_i} EU_{\theta_i^*}^{\mu^*}(m_i)$ .

By the strict concavity assumption, best responses are unique and minimize the absolute value of the derivative of the expected utility function. We examine the difference in the first-order conditions at the optimum for  $\mu$  and  $\mu^*$ , for the case in which  $m_i^* > \theta_i^*$  (the proof is analogous for  $m_i^* < \theta_i^*$ ):  $\frac{\partial EU_{\theta_i^*}^\mu}{\partial m_i}(m_i^*) - \frac{\partial EU_{\theta_i^*}^{\mu^*}}{\partial m_i}(\theta_i^*)$ , where for any  $\theta_i$ ,  $m_i$ , and  $\mu$ ,

$$(B1) \quad \frac{\partial EU_{\theta_i}^\mu}{\partial m_i}(m_i) = \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(m_i, m_{-i}; \theta_i, \theta_{-i}) d\mu.$$

Since, by assumption,  $EU_{\theta_i^*}^\mu(m_i)$  is strictly concave and maximized at  $m_i^*$ , whereas  $EU_{\theta_i^*}^{\mu^*}(m_i)$  is strictly concave and maximized at  $\theta_i^*$ , if  $m_i^* > \theta_i^*$  it follows that  $\frac{\partial EU_{\theta_i^*}^\mu}{\partial m_i}(m_i^*) - \frac{\partial EU_{\theta_i^*}^{\mu^*}}{\partial m_i}(\theta_i^*) \geq 0$ . Using (B1), this can be rewritten as

$$\int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(m_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu - \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu^* \geq 0.$$

Next, we add and subtract  $\int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu$ , and rearrange terms to obtain

$$\begin{aligned} \mathbb{A}_i &:= \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu - \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(m_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu \\ &\leq \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu - \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu^* =: \mathbb{B}_i. \end{aligned}$$

By the mean value theorem, there exists  $m'_i \in [\theta_i^*, m_i^*]$  such that

$$\mathbb{A}_i = \left( \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial^2 m_i}(m'_i, m_{-i}; \theta_i^*, \theta_{-i}) d\mu \right) \cdot (\theta_i^* - m_i^*).$$

Since  $l = (m_i^* - \theta_i^*) > 0$ , and expected payoffs are strictly concave, this can be written as

$$\mathbb{A}_i = \left| \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial^2 m_i}(m'_i, m_{-i}; \theta_i^*, \theta_{-i}) \cdot d\mu \right| \cdot l.$$

Since  $\text{marg}_{\Theta_{-i}} \mu^* = \text{marg}_{\Theta_{-i}} \mu$  and  $\mu^* \in C_i^T$ , the term  $\mathbb{B}_i$  can be written as

$$\begin{aligned} \mathbb{B}_i &= \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) d\mu - \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i}(\theta_i^*, \theta_{-i}; \theta_i^*, \theta_{-i}) d\mu \\ &\leq \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \left( \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j}(\theta_i^*, m_{-i}; \theta_i^*, \theta_{-i}) \right| \cdot |\theta_j - m_j| \right) d\mu \\ &\leq \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j}(m_i, m_{-i}; \theta_i^*, \theta_{-i}) \right| d\mu \cdot l, \end{aligned}$$

where the first bound follows from the mean-value theorem (applied to  $\partial U_i / \partial m_i$  and  $m_{-i}$ ) and the triangle inequality, whereas the second bound follows from the maximality of  $l$ . Since, from above,  $\mathbb{A}_i \leq \mathbb{B}_i$ , we have that

$$\left| \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial^2 m_i}(m'_i, m_{-i}; \theta_i^*, \theta_{-i}) \cdot d\mu \right| \leq \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j}(m_i, m_{-i}; \theta_i^*, \theta_{-i}) \right| d\mu,$$

which contradicts the  $\mathcal{B}$ -LSE condition for  $i$ . ■

#### PROOF OF LEMMA 1:

If the direct mechanism  $(d, t)$  is ex post incentive compatible, then the first-order conditions, which guarantee that truthful revelation is an ex post equilibrium,

imply that  $\left( \frac{\partial v_i}{\partial x}(d(m), \theta) \cdot \frac{\partial d}{\partial \theta_i}(m) + \frac{\partial t_i}{\partial m_i}(m) \right) \Big|_{m=\theta} = 0$  for all  $i$  and  $\theta$ . Hence,  $\frac{\partial t_i}{\partial m_i}(\theta) = -\frac{\partial v_i}{\partial x}(d(\theta), \theta) \cdot \frac{\partial d}{\partial \theta_i}(\theta)$  for all  $\theta$ . Integrating over  $m_i$ , it follows that for every  $m$ ,

$$(B2) \quad t_i(m_i, m_{-i}) = -\int_{\underline{\theta}_i}^{m_i} \frac{\partial v_i}{\partial x}(d(s, m_{-i}), s, m_{-i}) \cdot \frac{\partial d}{\partial \theta_i}(s, m_{-i}) ds + K(m_{-i}).$$

For every  $i$ , let  $\varpi_i : \Theta \rightarrow \mathbb{R}$  be such that  $\varpi_i(\theta_i, \theta_{-i}) = v_i(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})$ , and notice that

$$\frac{\partial \varpi_i}{\partial \theta_i}(\theta_i, \theta_{-i}) = \frac{\partial v_i}{\partial x}(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \cdot \frac{\partial d}{\partial \theta_i}(\theta_i, \theta_{-i}) + \frac{\partial v_i}{\partial \theta_i}(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}).$$

Thus, (B2) can be rewritten as

$$\begin{aligned} t_i(m_i, m_{-i}) &= -\int_{\underline{\theta}_i}^{m_i} \frac{\partial \varpi_i}{\partial \theta_i}(s, m_{-i}) ds + \int_{\underline{\theta}_i}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s, m_{-i}), s, m_{-i}) ds + K(m_{-i}) \\ &= -v_i(d(m_i, m_{-i}), m_i, m_{-i}) \\ &\quad + \int_{\underline{\theta}_i}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s, m_{-i}), s, m_{-i}) ds + K(m_{-i}) + v_i(d(\underline{\theta}_i, m_{-i}), \underline{\theta}_i, m_{-i}). \end{aligned}$$

Recall the canonical transfers  $t^*$  as in (8) and notice that the result follows by letting  $\tau_{-i}(m_{-i}) = K(m_{-i}) + v_i(d(\underline{\theta}_i, m_{-i}), \underline{\theta}_i, m_{-i})$ . ■

#### PROOF OF PROPOSITION 1:

The “if” part is immediate. For the “only if,” suppose that  $d$  is belief-free implemented by a direct mechanism with transfer scheme  $t$ . Then, the direct mechanism  $(d, t)$  is strictly ex post incentive compatible and Lemma 1 implies that  $t_i(m) = t_i^*(m) + \tau_i(m_{-i})$  for some  $\tau_i : M_{-i} \rightarrow \mathbb{R}$ . Hence,  $(d, t)$  and the canonical direct mechanism  $(d, t^*)$  induce the same ex post best responses, and hence the same sets of belief-free Rationalizable strategies. The canonical direct mechanism therefore also belief-free implements  $d$ . ■

#### PROOF OF THEOREM 2:

The proof is as explained in the main text: with transfers as in (10), truthful revelation satisfies the first-order conditions for any  $\mu \in C_i^T$ . Condition 1 in Theorem 2 ensures the strict concavity condition in Theorem 1, and so the second-order conditions are also satisfied. The mechanism therefore is  $\mathcal{B}$ -IC. Condition 2 in the Theorem implies the  $\mathcal{B}$ -LSE Condition of Theorem 1. The result thus follows from Theorem 1. ■

## PROOF OF LEMMA 2:

Notice that in the canonical direct mechanism,

$$(B3) \quad \frac{\partial U_i}{\partial m_i}(m; \theta) = \left( \frac{\partial v_i}{\partial x}(d(m), \theta) - \frac{\partial v_i}{\partial x}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) =: W_i(m; \theta).$$

Hence, if opponents are truthful, then truthful revelation satisfies the necessary first-order conditions: that is,  $W_i(\theta; \theta) = 0$ . Furthermore, the second order derivative is

$$(B4) \quad \begin{aligned} \frac{\partial^2 U_i}{\partial^2 m_i}(m; \theta) &= \frac{\partial W_i}{\partial m_i}(m; \theta) \\ &= \left( \left( \frac{\partial^2 v_i}{\partial^2 x}(d(m), \theta) - \frac{\partial^2 v_i}{\partial^2 x}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) - \frac{\partial^2 v_i}{\partial x \partial \theta_i}(d(m), m) \right) \frac{\partial d}{\partial \theta_i}(m) \\ &\quad + \left( \frac{\partial v_i}{\partial x}(d(m), \theta) - \frac{\partial v_i}{\partial x}(d(m), m) \right) \frac{\partial^2 d}{\partial^2 \theta_i}(m), \end{aligned}$$

which at  $m = \theta$  simplifies to

$$\frac{\partial^2 U_i}{\partial^2 m_i}(\theta; \theta) = - \frac{\partial^2 v_i}{\partial x \partial \theta_i}(d(\theta), \theta) \frac{\partial d}{\partial \theta_i}(\theta).$$

Under the single crossing condition as in Assumption 1,  $\frac{\partial^2 v_i}{\partial x \partial \theta_i}(d(\theta), \theta) > 0$ , hence the necessary second-order conditions for a local maximum are satisfied only if  $\frac{\partial d}{\partial \theta_i}(\theta) \geq 0$ .

If  $(d, t^*)$  is strictly EPIC, then  $\frac{\partial d}{\partial \theta_i}(\theta) \geq 0$  must hold for all  $i$  and  $\theta_i$ . However,  $\frac{\partial d}{\partial \theta_i}(\theta) = 0$  can only hold for isolated points. To show this, consider that  $\frac{\partial d}{\partial \theta_i}(s_i, \theta_{-i}) = 0$  for some interval  $s_i \in (\theta_i, \theta_i + \varepsilon)$ , where  $\varepsilon > 0$ . Then, the FOC  $W_i(\theta; \theta) = W_i(s_i, \theta_{-i}; \theta) = 0$  and all  $s_i \in (\theta_i, \theta_i + \varepsilon)$  ensure the same utility as reporting the true type  $\theta_i$ , which contradicts strict EPIC. Hence,  $d$  can not be constant on an interval, therefore  $d$  is strictly increasing in all  $\theta_i$ .

In the other direction, if  $d$  is strictly increasing, a report  $m_i$  can satisfy the FOC given  $\theta_i$  and  $m_{-i} = \theta_{-i}$ ,  $W_i(m_i, \theta_{-i}; \theta) = 0$ , if either term in (B3) is 0. By the single crossing condition, the first term is 0 if and only if  $m_i = \theta_i$ . The second term can be 0 for some  $m_i \neq \theta_i$  if  $\frac{\partial d}{\partial \theta_i}(m_i, \theta_{-i}) = 0$ . Hence, the FOC for  $i$ 's optimization problem, when the state is  $\theta$  and the opponents report truthfully, can only be satisfied by (i)  $m_i = \theta_i$  and by (ii)  $m_i \neq \theta_i$  such that  $\frac{\partial d}{\partial \theta_i}(m_i, \theta_{-i}) = 0$ . We show next that only the first case satisfies the second-order conditions for a local optimum.

**Case (i):** Consider  $m_i = \theta_i$ . Then, either  $\frac{\partial^2 U_i}{\partial^2 m_i}(\theta; \theta) < 0$  (which implies that truthful revelation is a strict local optimum), or  $\frac{\partial d}{\partial \theta_i}(\theta) = 0$ . But since  $d$  is strictly increasing, it may have a zero derivative only at an isolated point. Hence, there

exists a neighborhood around  $\theta_i$  such that for all  $s_i \neq \theta_i$  in this neighborhood,  $\frac{\partial d}{\partial \theta_i}(s_i, \theta_{-i}) > 0$ . This in turn implies that  $U_i(\cdot, \theta_{-i}; \theta)$  is strictly concave around  $\theta_i$ , and hence truthful revelation is a strict local optimum.

**Case (ii):** Suppose that  $m_i \neq \theta_i$  is such that  $\frac{\partial d}{\partial \theta_i}(m_i, \theta_{-i}) = 0$ . Then, by (B4),  $\frac{\partial^2 U_i}{\partial^2 m_i}(m_i, \theta_{-i}; \theta) = \left( \frac{\partial v_i}{\partial x}(d(m_i, \theta_{-i}), \theta) - \frac{\partial v_i}{\partial x}(d(m_i, \theta_{-i}), m_i, \theta_{-i}) \right) \frac{\partial^2 d}{\partial^2 \theta_i}(m_i, \theta_{-i})$ . By the single crossing condition, the term in parentheses has the same sign around  $m_i$  in a neighborhood away from  $\theta_i$ . If  $\frac{\partial d}{\partial \theta_i}(m_i, \theta_{-i}) = 0$ , the differentiability and strict monotonicity of  $d$  implies that  $d$  switches convexity at  $m_i$ , and its second-order derivative is negative to the left of  $m_i$ , and positive to the right of  $m_i$ . It follows that  $\frac{\partial^2 U_i}{\partial^2 m_i}(\cdot, \theta_{-i}; \theta)$  switches sign at  $m_i$ , and hence  $m_i \neq \theta_i$  is not optimal if  $\frac{\partial d}{\partial \theta_i}(m_i, \theta_{-i}) = 0$ .

Given that the FOC for an interior optimum can only be satisfied at (i) and (ii), to show that (i) is the global optimum, it suffices to show that the end points  $\underline{\theta}_i$  and  $\bar{\theta}_i$  are not optimal for interior types. The single crossing condition and the strict monotonicity of  $d$  imply that the right-derivative of  $W_i(\underline{\theta}_i, \theta_{-i}; \theta)$  is positive, thus  $\underline{\theta}_i$  is not optimal, and that the left-derivative of  $W_i(\bar{\theta}_i, \theta_{-i}; \theta)$  is negative, thus  $\bar{\theta}_i$  is not optimal. Strict EPIC follows. ■

#### PROOF OF PROPOSITION 2:

By Lemma 2, in SCC-environments the canonical direct mechanism is strictly ex post incentive compatible, and if  $\frac{\partial W_i}{\partial m_i}(m; \theta) < 0$ , it also satisfies the strict concavity condition of Theorem 1 for  $\mathcal{B} = \mathcal{B}^{BF}$ . The result then follows directly from Corollary 1. ■

#### PROOF OF PROPOSITION 3:

**Step 1:** Note that in SCC-PC-environments, for any  $i$  and  $j$ , we have that  $\frac{\partial W_i}{\partial m_j}(m; \theta) = -\frac{\partial^2 v_i}{\partial x \partial \theta_j}(d(m), m) \frac{\partial d}{\partial \theta_i}(m)$  and in particular  $\frac{\partial W_i}{\partial m_i}(m; \theta) < 0$  for all  $m, \theta$ . For  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$  defined as in (13), the independent common prior implies that, for  $\hat{f}_i(\theta_i) := E(\hat{L}_i(\theta_{-i}) | \theta_i)$ , we have  $\hat{f}_i'(\theta_i) = 0$  for all  $\theta_i$ . Hence, Condition 1 of Theorem 2 holds.

**Step 2:** By construction,  $\frac{\partial W_i}{\partial m_j}(m; \theta) = -\frac{\partial \hat{L}_i}{\partial m_j}$ . Hence, the RHS of Condition 2 of Theorem 2 is 0 for all  $\theta_i$ , and hence smaller than the LHS (which therefore satisfies Condition 2). Since both conditions of Theorem 2 are satisfied, the transfers in (14) ensure full implementation. Moreover,  $\frac{\partial W_i}{\partial m_j}(m; \theta) = -\frac{\partial \hat{L}_i}{\partial m_j}$  also ensures that  $i$ 's expected payoffs in the modified mechanism, and hence the set of best replies, are constant in  $m_j$  for all  $j \neq i$ . But since  $\theta_i$  is the unique best reply to all conjectures that assign probability one to others' truthful reports, it also follows that  $R_{\theta_i}^{\mathcal{B}, 1} = \{\theta_i\}$  for all  $\theta_i$  and  $i$ . Hence, full implementation is achieved in dominant strategies. ■

## PROOF OF PROPOSITION 4:

As noted in the previous proof, we have that  $\frac{\partial W_i}{\partial m_j}(m; \theta) = -\frac{\partial^2 v_i}{\partial x \partial \theta_j}(d(m), m) \frac{\partial d}{\partial \theta_i}(m)$  for all  $i$  and  $j$  in SCC-PC-environments. For  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$  defined as in (13), and  $\hat{f}_i(\theta_i) := E(\hat{L}_i(\theta_{-i}) | \theta_i)$ , Theorem 5 in Milgrom and Weber (1982) implies that, if valuations are supermodular, then  $\hat{f}'_i(\theta_i) \geq 0$  for all  $\theta_i$ . Since, by the assumptions of SCC-environments  $\frac{\partial W_i}{\partial m_i}(m; \theta) < 0$ , it follows that  $\frac{\partial W_i}{\partial m_i}(m; \theta) - f'_i(m_i) < 0$  for all  $m, \theta$ , which implies that Condition 1 of Theorem 2 holds. The rest is identical to Step-2 of the proof of Proposition 3. ■

## PROOF OF PROPOSITION 5:

If  $(d, t^*)$  is strictly EPIC, then  $d$  is strictly increasing (by Lemma 2) and  $\mathcal{B}$ -DS implementability follows from Propositions 3 and 4. For the other direction, if  $d$  is  $\mathcal{B}$ -DS implementable, then there exist transfers  $t^{DS}$  that guarantee that truthful revelation is the only interim best response, regardless of the opponents' strategies. Hence, for all  $\theta_i$ ,

$$\{\theta_i\} = \arg \max_{\theta'_i \in M_i} \int_{\Theta_{-i}} U_i(\theta'_i, m_{-i}; \theta_i, \theta_{-i}) db_{\theta_i} \quad \text{for all } m_{-i},$$

where  $b_{\theta_i}$  is such that  $B_{\theta_i} = \{b_{\theta_i}\}$  from the common prior assumption. The necessary condition for this best response is that for all  $i$ , for all  $\theta_i$ , and for all  $m_{-i} \in M_{-i}$ ,

$$\left( \int_{\Theta_{-i}} \left( \frac{\partial v_i}{\partial x}(d(m), \theta) \cdot \frac{\partial d}{\partial \theta_i}(m) \right) db_{\theta_i} + \frac{\partial t_i^{DS}}{\partial m_i}(m_i, m_{-i}) \right) \Big|_{m_i = \theta_i} = 0, \quad \text{hence}$$

$$\frac{\partial t_i^{DS}}{\partial m_i}(\theta_i, m_{-i}) = - \int_{\Theta_{-i}} \frac{\partial v_i}{\partial x}(d(\theta_i, m_{-i}), \theta_i, \theta_{-i}) db_{\theta_i} \cdot \frac{\partial d}{\partial \theta_i}(\theta_i, m_{-i}).$$

Using this property of the transfers, the second order partial derivative of the interim payoffs is the expected value of (B4), which by Assumption 2 at  $m_i = \theta_i$  simplifies to

$$\int_{\Theta_{-i}} \frac{\partial^2 U_i}{\partial^2 m_i}(\theta_i, m_{-i}; \theta) db_{\theta_i} = - \frac{\partial d}{\partial \theta_i}(\theta_i, m_{-i}) \cdot \int_{\Theta_{-i}} \left( \frac{\partial^2 v_i}{\partial x \partial \theta_i}(d(\theta_i, m_{-i}), \theta_i, m_{-i}) \right) db_{\theta_i}.$$

By the single crossing condition in Assumption 1, the second-order conditions are satisfied only if  $\frac{\partial d}{\partial \theta_i}(m) \geq 0$ . However,  $\frac{\partial d}{\partial \theta_i}(\theta) = 0$  can only hold for isolated points, and so  $d$  must be strictly increasing. To show this, suppose (by means of contradiction) that there exists  $\theta_i$  and  $\varepsilon > 0$  and some  $s_{-i}$  such that  $\frac{\partial d}{\partial \theta_i}(s_i, s_{-i}) = 0$  for all  $s_i \in (\theta_i, \theta_i + \varepsilon)$ , which also implies for  $W_i$  (as given by (B3)) that  $W_i(s_i, s_{-i}; \theta) = 0$  for all  $s_i \in (\theta_i, \theta_i + \varepsilon)$ . Thus, for the FOC,  $\int_{\Theta_{-i}} W_i(\theta_i, s_{-i}; \theta) db_{\theta_i} = \int_{\Theta_{-i}} W_i(s_i, s_{-i}; \theta) db_{\theta_i} = 0$  for all  $s_i \in (\theta_i, \theta_i + \varepsilon)$ , in other words, any such  $s_i$  ensures the same expected payoff as reporting the true type  $\theta_i$ , hence contradicting  $\mathcal{B}$ -DS implementation. Hence,  $d$  is strictly increasing for all  $\theta_i$  and hence  $(d, t^*)$  is strictly EPIC (by Lemma 2). ■



## PROOF OF PROPOSITION 6:

For each  $i, j$ , let  $\varphi_{ji} : \Theta_i \rightarrow \Theta_j$  be such that, for each  $\theta_i \in \Theta_i$ ,  $\varphi_{ji}(\theta_i) := E(\theta_j | \theta_i)$ . By assumption, the functions  $\varphi_{ji}$  are differentiable. Then, the designer's information is represented by belief restrictions  $\bar{\mathcal{B}} = ((\bar{\mathcal{B}}_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$  such that  $\bar{\mathcal{B}}_{\theta_i} = \{\beta \in \Delta(\Theta_{-i}) : E_{\beta}(\theta_j) = \varphi_{ji}(\theta_i) \text{ for all } j \in I \setminus \{i\}\}$ , for each  $i \in I$  and  $\theta_i \in \Theta_i$ . Next notice that, in SCC-PC-environments, the function  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$  defined in (13) is linear. Hence, if the conditional expectations  $E(\theta_j | \theta_i)$  are common knowledge in  $\bar{\mathcal{B}}$ , so are the conditional expectations  $E(\hat{L}_i(\theta_{-i}) | \theta_i)$ , which are thus *moment conditions* that can be used to weaken the strategic externalities. Formally, let  $\hat{f}_i(\theta_i) := \hat{L}_i((\varphi_{ji}(\theta_i))_{j \in I \setminus \{i\}})$ . Then, because of the linearity of the  $E(\cdot)$  operator and of  $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$ , we have that  $E(\hat{L}_i(\theta_{-i}) | \theta_i) = \hat{f}_i(\theta_i)$  for all  $i$ , that is  $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I} \in \varrho(\bar{\mathcal{B}})$ . Moreover,  $\hat{f}_i$  is nondecreasing if so are the functions  $\varphi_{ji}$ . The result then follows from Theorem 2 for the same reasons as Proposition 4 does. ■

## PROOF OF THEOREM 3:

Fix  $i$  and  $\theta_i$ . Let  $l := \max_{k \in I} l_k$ . By the definition of  $t^\rho$ , for any  $\mu \in C_{\theta_i}^{\mathcal{B}}$ , adding and subtracting  $L_i(\theta_{-i})$ , applying Leibniz's rule and the triangle inequality, we have

$$\begin{aligned} \text{(B5)} \quad \left| \frac{\partial EU_{\theta_i}^\mu}{\partial m_i}(\theta_i) \right| &= \left| \int_{\Theta_{-i} \times M_{-i}} \left( \frac{\partial v_i}{\partial x}(d(\theta_i, m_{-i}), \theta) - \frac{\partial v_i}{\partial x}(d(\theta_i, m_{-i}), \theta_i, m_{-i}) \right) \right. \\ &\quad \times \left. \frac{\partial d}{\partial \theta_i}(\theta_i, m_{-i}) + L_i(m_{-i}) - L_i(\theta_{-i}) + L_i(\theta_{-i}) - f_i(\theta_i) d\mu \right| \\ &\leq \int_{\Theta_{-i} \times M_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j}(\theta; \theta_i, s_{-i}) \right| |\theta_j - m_j| d\mu + \varepsilon \leq SE_i^\rho(\theta_i) \cdot l + \varepsilon. \end{aligned}$$

For any  $m_i \in R_i^{\mathcal{B},1}(\theta_i)$ , there exists  $\mu \in C_{\theta_i}^{\mathcal{B}}$  such that  $m_i \in BR_{\theta_i}(\mu)$ . Since  $m_i$  is best reply, it minimizes the first-order partial derivative. Using (B5) and by the concavity of the expected utility function, it follows that for all  $\mu \in C_{\theta_i}^{\mathcal{B}}$ ,  $\left| \frac{\partial EU_{\theta_i}^\mu}{\partial m_i}(m_i) - \frac{\partial EU_{\theta_i}^\mu}{\partial m_i}(\theta_i) \right| \leq SE_i^\rho(\theta_i) \cdot l + \varepsilon$ . By the mean value theorem, there exists  $s_i \in M_i$  such that  $\left| \left( \partial^2 EU_{\theta_i}^\mu / \partial^2 m_i \right)(s_i) \right| |m_i - \theta_i| \leq SE_i^\rho(\theta_i) \cdot l + \varepsilon$ . Therefore, for all  $\theta_i$  and  $m_i \in R_i^{\mathcal{B},1}(\theta_i)$ ,

$$\text{(B6)} \quad |m_i - \theta_i| \leq \frac{SE_i^\rho(\theta_i) \cdot l + \varepsilon}{OC_i^\rho(\theta_i)}.$$

Then, for any  $m_i \in R_i^{\mathcal{B},2}(\theta_i)$ , there exists  $\mu \in C_{\theta_i}^{\mathcal{B}} \cap R_i^{\mathcal{B},1}(\theta_i)$  such that  $m_i^* \in BR_{\theta_i}(\mu^*)$ . For the Taylor-expansion of  $\partial EU_{\theta_i}^\mu / \partial m_i$  at  $\theta_i$  around  $m_i$  there exists  $s_i \in M_i$  such that

$$\frac{\partial EU_{\theta_i}^\mu}{\partial m_i}(\theta_i) = \frac{\partial EU_{\theta_i}^\mu}{\partial m_i}(m_i) + \frac{\partial^2 EU_{\theta_i}^\mu}{\partial^2 m_i}(s_i) (\theta_i - m_i).$$

Since  $m_i$  is best reply to  $\mu$  and  $EU_{\theta_i}^\mu(m_i)$  is strictly concave, we have that

$$(B7) \quad \left| \frac{\partial^2 EU_{\theta_i}^\mu(s_i)}{\partial^2 m_i} \right| |\theta_i - m_i| \leq \left| \frac{\partial EU_{\theta_i}^\mu(\theta_i)}{\partial m_i} \right|.$$

Consider the RHS of (B7) and bound it similarly to (B5), by adding and subtracting  $L_i(\theta_{-i})$ , applying Leibniz's rule and the triangle inequality, to get

$$(B8) \quad \left| \frac{\partial EU_{\theta_i}^\mu(\theta_i)}{\partial m_i} \right| \leq \int_{\Theta_{-i} \times M_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j}(\theta; \theta_i, s_{-i}) \right| |\theta_j - m_j| d\mu + \varepsilon.$$

Let  $(h, \theta_h) := \arg \max_{i \in I, \theta_i \in \Theta_i} SE_i^\rho(\theta_i)/OC_i^\rho(\theta_i)$  and  $\widehat{SE}^\rho := SE_h^\rho(\theta_h)$ ,  $\widehat{OC}^\rho := OC_h^\rho(\theta_h)$ . Hence,  $NSE^\rho = \frac{\widehat{SE}^\rho}{\widehat{OC}^\rho}$ . Combining (B7), (B8), and (B6), we get

$$|\theta_i - m_i| \leq \frac{SE_i^\rho(\theta_i)}{OC_i^\rho(\theta_i)} \frac{\widehat{SE}^\rho \cdot l + \varepsilon}{\widehat{OC}^\rho} + \frac{\varepsilon}{OC_i^\rho(\theta_i)}$$

for all  $\theta_i$  and  $m_i \in R_i^{\mathcal{B},2}(\theta_i)$ . By induction, at the  $k$ th round, for all  $\theta_i$  and  $m_i \in R_i^{\mathcal{B},k}(\theta_i)$ ,

$$\begin{aligned} |\theta_i - m_i| &\leq \left( \frac{SE_i^\rho(\theta_i)}{OC_i^\rho(\theta_i)} \right) (NSE^\rho)^{k-1} \cdot l + \left( \frac{SE_i^\rho(\theta_i)}{OC_i^\rho(\theta_i)} \right) \sum_{n=0}^{k-2} (NSE^\rho)^n \\ &\quad \times \frac{\varepsilon}{\widehat{OC}^\rho} + \frac{\varepsilon}{OC_i^\rho(\theta_i)}. \end{aligned}$$

Condition 2 in Theorem 2 guarantees that  $NSE^\rho < 1$  and  $SE_i^\rho(\theta_i)/OC_i^\rho(\theta_i) < 1$ . Hence, taking limits as  $k \rightarrow \infty$  and letting  $OC^\rho := \min_{i \in I, \theta_i \in \Theta_i} OC_i^\rho(\theta_i)$ , for all  $\theta_i$  and  $m_i \in R_i^{\mathcal{B}}(\theta_i)$ :

$$\begin{aligned} |\theta_i - m_i| &\leq \frac{SE_i^\rho(\theta_i)}{OC_i^\rho(\theta_i)} \frac{1}{1 - NSE^\rho} \frac{\varepsilon}{\widehat{OC}^\rho} + \frac{\varepsilon}{OC_i^\rho(\theta_i)} \\ &\leq NSE^\rho \frac{1}{1 - NSE^\rho} \frac{\varepsilon}{OC^\rho} + \frac{\varepsilon}{OC^\rho} = \frac{1}{1 - NSE^\rho} \frac{\varepsilon}{OC^\rho}. \end{aligned}$$

Hence, for all  $i$  and  $\theta_i$ ,  $R_i^{\mathcal{B}}(\theta_i) \subseteq [\theta_i \pm \varepsilon / ((1 - NSE^\rho) OC^\rho)]$ . ■

## REFERENCES

- Aradillas-López, Andrés, Amit Gandhi, and Daniel Quint. 2013. "Identification and Inference in Ascending Auctions with Correlated Private Values." *Econometrica* 81 (2): 489–534.
- Arieli, Itai. 2010. "Rationalizability in Continuous Games." *Journal of Mathematical Economics* 46 (5): 912–24.
- Artemov, Georgy, Takashi Kunimoto, and Roberto Serrano. 2013. "Robust Virtual Implementation: Toward a Reinterpretation of the Wilson Doctrine." *Journal of Economic Theory* 148 (2): 424–47.

- Ausubel, Lawrence M., and Oleg V. Baranov. 2013. "Core-Selecting Auctions with Incomplete Information." <http://www.obaranov.com/docs/Ausubel-Baranov-Core-Selecting-Auctions-with-Incomplete-Information.pdf>.
- Battigalli, Pierpaolo. 2003. "Rationalizability in Infinite, Dynamic Games with Incomplete Information." *Research in Economics* 57 (1): 1–38.
- Battigalli, Pierpaolo, Alfredo Di Tillio, Edoardo Grillo, and Antonio Penta. 2011. "Interactive Epistemology and Solution Concepts for Games with Asymmetric Information." *B.E. Journal of Theoretical Economics: Advances in Theoretical Economics* 11 (1).
- Battigalli, Pierpaolo, and Marciano Siniscalchi. 2003. "Rationalization and Incomplete Information." *Advances in Theoretical Economics* 3 (1).
- Bergemann, Dirk, and Stephen Morris. 2005. "Robust Mechanism Design." *Econometrica* 73 (6): 1771–1813.
- Bergemann, Dirk, and Stephen Morris. 2007. "An Ascending Auctions for Independent Values: Uniqueness and Robustness to Strategic Uncertainty." *American Economic Review* 97 (2): 125–30.
- Bergemann, Dirk, and Stephen Morris. 2009a. "Robust Implementation in Direct Mechanisms." *Review of Economic Studies* 76 (4): 1175–1204.
- Bergemann, Dirk, and Stephen Morris. 2009b. "Robust Virtual Implementation." *Theoretical Economics* 4 (1): 45–88.
- Bergemann, Dirk, and Stephen Morris. 2011. "Robust Implementation in General Mechanisms." *Games and Economic Behavior* 71 (2): 261–81.
- Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman. 2015. "Linear Social Interactions Models." *Journal of Political Economy* 123 (2): 444–96.
- Borgers, Tilman, and Doug Smith. 2012. "Robustly Ranking Mechanisms." *American Economic Review* 102 (3): 325–29.
- Borgers, Tilman, and Doug Smith. 2014. "Robust Mechanism Design and Dominant Strategy Voting Rules." *Theoretical Economics* 9 (2): 339–60.
- Carroll, Gabriel. 2015. "Robustness and Linear Contracts." *American Economic Review* 105 (2): 536–63.
- Chung, Kim-Sau, and Jeffrey C. Ely. 2001. "Efficient and Dominance Solvable Auctions with Interdependent Values." <https://sites.google.com/site/kimsauchung/dominance.pdf>.
- Cr mer, Jacques, and Richard P. McLean. 1985. "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent." *Econometrica* 53 (2): 345–61.
- Cr mer, Jacques, and Richard P. McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* 56 (6): 1247–57.
- Dasgupta, Partha, and Eric S. Maskin. 2000. "Efficient Auctions." *Quarterly Journal of Economics* 115 (2): 341–88.
- d'Aspremont, Claude, Jacques Cr mer, and Louis-Andr  G rard-Varet. 1998. "Unique Implementation in Auctions and in Public Goods Problems." CORE Discussion Paper 1998010.
- d'Aspremont, Claude, and Louis-Andr  G rard-Varet. 1979. "Incentives and Incomplete Information." *Journal of Public Economics* 11 (1): 25–45.
- Deb, Rahul, and Mallesh M. Pai. 2017. "Discrimination via Symmetric Auctions." *American Economic Journal: Microeconomics* 9 (1): 275–314.
- Dekel, Eddie, Drew Fudenberg, and Stephen Morris. 2007. "Interim Correlated Rationalizability." *Theoretical Economics* 2 (1): 15–40.
- Duggan, John, and Joanne Roberts. 2002. "Implementing the Efficient Allocation of Pollution." *American Economic Review* 92 (4): 1070–78.
- Fainmesser, Itay P., and Andrea Galeotti. 2015. "Pricing Network Effects." *Review of Economic Studies* 83 (1): 165–98.
- Gershkov, Alex, Jacob K. Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. 2013. "On the Equivalence of Bayesian and Dominant Strategy Implementation." *Econometrica* 81 (1): 197–220.
- Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70 (3): 393–408.
- Guo, Huiyi, and Nicholas C. Yannellis. 2017. "Robust Strong Nash Implementation." Unpublished.
- Healy, Paul J., and Laurent Mathevet. 2012. "Designing Stable Mechanisms for Economic Environments." *Theoretical Economics* 7 (3): 609–61.
- Hellwig, Martin F. 1980. "On the Aggregation of Information in Competitive Markets." *Journal of Economic Theory* 22 (3): 477–98.
- Jackson, Matthew O. 1991. "Bayesian Implementation." *Econometrica* 59 (2): 461–77.
- Jackson, Matthew O. 1992. "Implementation in Undominated Strategies: A Look at Bounded Mechanisms." *Review of Economic Studies* 59 (4): 757–75.
- Jehiel, Philippe, Moritz Meyer-ter-Vehn, and Benny Moldovanu. 2012. "Locally Robust Implementation and Its Limits." *Journal of Economic Theory* 147 (6): 2439–52.
- Jehiel, Philippe, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William R. Zame. 2006. "The Limits of Ex Post Implementation." *Econometrica* 74 (3): 585–610.

- Kos, Nenad, and Matthias Messner. 2015. "Selling to the Mean." Unpublished.
- Laffont, Jean-Jacques, and Eric Maskin. 1980. "A Differential Approach to Dominant Strategy Mechanisms." *Econometrica* 48 (6): 1507–20.
- Li, Yunan. 2016. "Approximation in Mechanism Design with Interdependent Values." *Games and Economic Behavior*. <https://doi.org/10.1016/j.geb.2016.01.006>.
- Lopomo, Giuseppe, Luca Rigotti, and Chris Shannon. 2014. "Uncertainty in Mechanism Design." <http://www.pitt.edu/~luca/Papers/mechanismdesign.pdf>.
- Manelli, Alejandro M., and Daniel R. Vincent. 2010. "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model." *Econometrica* 78 (6): 1905–38.
- Maskin, Eric. 1999. "Nash Equilibrium and Welfare Optimality." *Review of Economic Studies* 66 (1): 23–38.
- Mathevet, Laurent. 2010. "Supermodular Mechanism Design." *Theoretical Economics* 5 (3): 403–43.
- Mathevet, Laurent, and Ina Taneva. 2013. "Finite Supermodular Design with Interdependent Valuations." *Games and Economic Behavior* 82: 327–49.
- McLean, Richard, and Andrew Postlewaite. 2002. "Informational Size and Incentive Compatibility." *Econometrica* 70 (6): 2421–53.
- McLean, Richard, and Andrew Postlewaite. 2004. "Informational Size and Efficient Auctions." *Review of Economic Studies* 71 (3): 809–27.
- Milgrom, Paul R., and Robert J. Weber. 1982. "A Theory of Auctions and Competitive Bidding." *Econometrica* 50 (5): 1089–1122.
- Mookherjee, Dilip, and Stefan Reichelstein. 1992. "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules." *Journal of Economic Theory* 56 (2): 378–99.
- Moulin, Herve. 1984. "Dominance Solvability and Cournot Stability." *Mathematical Social Sciences* 7 (1): 83–102.
- Müller, Christoph. 2016. "Robust Virtual Implementation under Common Strong Belief in Rationality." *Journal of Economic Theory* 162: 407–50.
- Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1): 58–73.
- Ollár, Mariann, and Antonio Penta. 2016. "Rationalizable Implementation with Belief Restrictions." Unpublished.
- Ollár, Mariann, and Antonio Penta. 2017. "Full Implementation via Transfers with Identical but Unknown Distributions." Unpublished.
- Oury, Marion, and Olivier Tercieux. 2012. "Continuous Implementation." *Econometrica* 80 (4): 1605–37.
- Penta, Antonio. 2011. "Backward Induction Reasoning in Incomplete Information Games." Unpublished.
- Penta, Antonio. 2012. "Higher Order Uncertainty and Information: Static and Dynamic Games." *Econometrica* 80 (2): 631–60.
- Penta, Antonio. 2013. "On the Structure of Rationalizability for Arbitrary Spaces of Uncertainty." *Theoretical Economics* 8 (2): 405–30.
- Penta, Antonio. 2015. "Robust Dynamic Implementation." *Journal of Economic Theory* 160: 280–316.
- Postlewaite, Andrew, and David Schmeidler. 1986. "Implementation in Differential Information Economies." *Journal of Economic Theory* 39 (1): 14–33.
- Roughgarden, Tim, and Inbal Talgam-Cohen. 2013. "Optimal and Near-Optimal Mechanism Design with Interdependent Values." In *Proceedings of the 14th ACM Conference on Electronic Commerce*, 767–84. New York: ACM.
- Vives, Xavier. 2011. "Strategic Supply Function Competition with Private Information." *Econometrica* 79 (6): 1919–66.
- Weinstein, Jonathan, and Muhamet Yildiz. 2007. "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements." *Econometrica* 75 (2): 365–400.
- Weinstein, Jonathan, and Muhamet Yildiz. 2011. "Sensitivity of Equilibrium Behavior to Higher-Order Beliefs in Nice Games." *Games and Economic Behavior* 72 (1): 288–300.
- Weinstein, Jonathan, and Muhamet Yildiz. 2013. "Robust Predictions in Infinite-Horizon Games: An Unrefinable Folk Theorem." *Review of Economic Studies* 80 (1): 365–94.
- Wilson, Robert. 1979. "Auctions of Shares." *Quarterly Journal of Economics* 93 (4): 675–89.
- Wilson, Robert B. 1987. "Game-Theoretic Analysis of Trading Processes." In *Advances in Economic Theory: Fifth World Congress*, edited by Truman F. Bewley, 33–70. New York: Cambridge University Press.
- Wolitzky, Alexander. 2016. "Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade." *Theoretical Economics* 11 (3): 971–1004.
- Yamashita, Takuro. 2015. "Implementation in Weakly Undominated Strategies: Optimality of Second-Price Auction and Posted-Price Mechanism." *Review of Economic Studies* 82 (3): 1223–46.